Impact of Severe Weather Events (USA)

Thursday, December 26, 2015

### Severe weather events impacting the economic and population health across the United States from 1950 to 2011.

### Synopis

The aim of this report is to explore and identify which major weather events across the United States are most harmful with respect to population health, i.e. fatalities and injuries; and which types of events have the greatest economic consequences with respect to property damage and crop damage. To identify these events, I explored and processed raw data from the U.S. National Oceanic and Atmospheric Administrations (NOAA) storm database. This database tracks and records characteristics of major weather events including the date, location and estimates of injuries, fatalities and damage incurred. Using data specifically from 1996 (the most complete and reliable records) I was able to identify the top weather events impacting popultion health with excessive heat responsible for most fatalities, and tornado events responsible for most injuries. With respect to economic impact, hurricane/typhoon events contributed most to the cost of damage.

### Loading and Data Processing

Data was obtained from the [Storm Events Database](https://www.ncdc.noaa.gov/stormevents/).

For this report we obtained files for the years [1950 to November 2011](https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2).

Documentation from [Storm Data Documentation](https://d396qusza40orc.cloudfront.net/repdata%2Fpeer2_doc%2Fpd01016005curr.pdf) was used to decipher variable names.

Supporting Weather event infomation was obtained from [FAQ](https://d396qusza40orc.cloudfront.net/repdata%2Fpeer2_doc%2FNCDC%20Storm%20Events-FAQ%20Page.pdf).

Storm Events Database Record updates and changes can be found [here](http://www.ncdc.noaa.gov/stormevents/details.jsp). This is of particular relevance to the data subset (from 1996 only) chosen for final analysis.

#### Downloading and Reading Data into R

The data for this report comes in the form of a comma-separated-value file compressed via the bzip2 algorithm to reduce its size. Note, for data processing and analysis it's assumed relevant packages are installed and loaded.

# Data download  
#if(!file.exists("./data")) {dir.create("./data")} #create a data dir if it doesn't exist  
#fileUrl <- 'https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2'  
#download.file(fileUrl, destfile='./data/stormData.csv.bz2') # Create file name and download dataset

library(dplyr)

##   
## Attaching package: 'dplyr'

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

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

library(knitr)

## Warning: package 'knitr' was built under R version 3.2.3

library(markdown)

## Warning: package 'markdown' was built under R version 3.2.2

#read data to R  
dataSet <- read.csv('./data/stormData.csv.bz2', sep=',', header=TRUE)

After reading the data in we check the dimensions, the first few rows and the variable types.

dim(dataSet)

## [1] 902297 37

#str(dataSet)  
head(dataSet,2)

## 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  
## EVTYPE BGN\_RANGE BGN\_AZI BGN\_LOCATI END\_DATE END\_TIME COUNTY\_END  
## 1 TORNADO 0 0  
## 2 TORNADO 0 0  
## COUNTYENDN END\_RANGE END\_AZI END\_LOCATI LENGTH WIDTH F MAG FATALITIES  
## 1 NA 0 14 100 3 0 0  
## 2 NA 0 2 150 2 0 0  
## INJURIES PROPDMG PROPDMGEXP CROPDMG CROPDMGEXP WFO STATEOFFIC ZONENAMES  
## 1 15 25.0 K 0   
## 2 0 2.5 K 0   
## LATITUDE LONGITUDE LATITUDE\_E LONGITUDE\_ REMARKS REFNUM  
## 1 3040 8812 3051 8806 1  
## 2 3042 8755 0 0 2

Looking at the documentation in the NOAA web site, specifically [here](http://www.ncdc.noaa.gov/stormevents/details.jsp), it highlights changes in the data collection and processing procedures over time. We can see that all event types are most complete (and defined) from 1996 to the present (48 from Directive 10-1605). Therefore, for this report we will explore data post 1996.

Before We can do this we will change the date format to allow subsetting of specific dates.

#First reformat the date column  
dataSet$BGN\_DATE <- strptime(dataSet$BGN\_DATE, '%m/%d/%Y %H:%M:%S')  
dataSet$BGN\_DATE <- as.Date(dataSet$BGN\_DATE)

And now we subset for events from 1996 onwards

#Filter for post 1996  
data96 <- subset(dataSet, BGN\_DATE >= '1996-1-1')

To assist exploratory analysis of the data we can filter for only variables of importance. This reduces the dataset from 902297 to 653530 entries.

#library(dplyr)  
# Select relevant variables  
stormData <- select(data96, BGN\_DATE, EVTYPE, FATALITIES, INJURIES, PROPDMG, PROPDMGEXP, CROPDMG, CROPDMGEXP)  
dim(stormData)

## [1] 653530 8

head(stormData,3)

## BGN\_DATE EVTYPE FATALITIES INJURIES PROPDMG PROPDMGEXP  
## 248768 1996-01-06 WINTER STORM 0 0 380 K  
## 248769 1996-01-11 TORNADO 0 0 100 K  
## 248770 1996-01-11 TSTM WIND 0 0 3 K  
## CROPDMG CROPDMGEXP  
## 248768 38 K  
## 248769 0   
## 248770 0

#### Event Type and Fatalities

For Fatalities we're interested in the 'FATALITIES' column and here is a summary. Highest single fatality event is 158 fatalities.

summary(stormData$FATALITIES)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.00000 0.00000 0.00000 0.01336 0.00000 158.00000

#Check for missing values  
any(is.na(stormData$FATALITIES))

## [1] FALSE

#### Event Type and Injuries

For Injuries we're interested in the 'INJURIES' column and here is a summary. Higest single Injury event is 1150 and there is no NAs.

summary(stormData$INJURIES) #Max is 1150 injuries

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.00e+00 0.00e+00 0.00e+00 8.87e-02 0.00e+00 1.15e+03

any(is.na(stormData$INJURIES))

## [1] FALSE

Checking the validity of the higher values we can see the largest single event, a tornado, caused 1150 injuries and also accounted for the largest Fatality event (identified in the Fatalities dataset above) of 158 occurring in Missouri, county of Jasper, on the 22-05-2011.

subset(data96, INJURIES > 1000)

## STATE\_\_ BGN\_DATE BGN\_TIME TIME\_ZONE COUNTY COUNTYNAME STATE  
## 862634 29 2011-05-22 04:40:00 PM CST 97 JASPER MO  
## EVTYPE BGN\_RANGE BGN\_AZI BGN\_LOCATI END\_DATE  
## 862634 TORNADO 2 SE CENTRAL CITY 5/22/2011 0:00:00  
## END\_TIME COUNTY\_END COUNTYENDN END\_RANGE END\_AZI END\_LOCATI  
## 862634 05:00:00 PM 0 NA 2 SSE DUENWEG  
## LENGTH WIDTH F MAG FATALITIES INJURIES PROPDMG PROPDMGEXP CROPDMG  
## 862634 9 1600 5 0 158 1150 2.8 B 0  
## CROPDMGEXP WFO STATEOFFIC  
## 862634 K SGF MISSOURI, Southwest  
## ZONENAMES  
## 862634   
## LATITUDE LONGITUDE LATITUDE\_E LONGITUDE\_  
## 862634 3703 9434 3703 9424  
## REMARKS  
## 862634 EPISODE NARRATIVE: A strong upper level trough across the central plains region and a strong upper level disturbance along with a frontal boundary in the area produced numerous severe storms with a few supercells across southeast Kansas. One cyclic supercell associated with the system produced a very large and deadly EF-5 tornado that hit the city of Joplin and then spawned two separate tornadoes that moved across southwestern Missouri. A separate tornadic supercell produced an EF-3 tornado that moved into McDonald County from Northeastern Oklahoma. \n\nContinuing thunderstorms lasted into July 23 as another line of thunderstorms moved across the region. This line of storms produced reports of large hail, damaging winds and a weak EF-0 tornado in south central Missouri. As the storms moved over the Joplin area, two police officers were struck by lightning while aiding in the post tornado efforts. One of the two officers later died of injuries from the lightning strike.EVENT NARRATIVE: National Weather Service survey teams rated the tornado that tracked across the southwest through east central portion of Joplin, Missouri, as an EF5 tornado. Maximum winds were estimated to have exceeded 200 miles per hour. The tornado had a maximum width of one mile and an overall path length of nearly 21.6 miles, nearly nine miles of which occurred in Jasper County.\n\nThe tornado killed 158 directly, three indirectly, and injured over 1150 people. Sadly, on May 24 a police officer who was volunteering from another department, was struck by lightning while serving in the response efforts and later died. Equally, a 56 year old man who had been included as a direct fatality was later determined to have died of a heart attack. Over 10,200 people filed for disaster assistance following the tornado.\n\nThe EF-5 rating (greater than 200 mph wind speeds) was mainly arrived at by the total destruction of vehicles, including some vehicles tossed several blocks and semi trucks thrown a quarter of a mile. Parking stops weighing over 300 pounds and re-barred into asphalt were uprooted and tossed. Other factors in the rating included damage to reinforced concrete structures, and that St. Johnb  
## REFNUM  
## 862634 862563

#### Event Type and Property Damage

For property damage we're interested in the 'PROPDMG' column. Here is a summary of property damage. There is also a multiplier column 'PROPDMGEXP' which may or may not increase these values.

summary(data96$PROPDMG) #Max = 5000

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.00 0.00 0.00 11.69 1.00 5000.00

any(is.na(data96$PROPDMG))

## [1] FALSE

Looking at the multipliers we can see there are quite a few levels and for this particular column there are K, M, B, 0, and entries with no multiplier. We can also see that there are no PROPDMG entries where there's no multiplier so we don't have to worry about missing some costs (entries that do not have a multiplier) when we filter for valid multipliers later.

unique(data96$PROPDMGEXP)

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

#Check for entries that have significant PROPDMG values but no Multiplier   
a <- subset(data96, PROPDMGEXP == '')  
summary(a$PROPDMG) #no values

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0 0 0 0 0 0

Looking closely at the observations with property damage greater equal to 5000 there appears to be double entry for "FLASH FLOOD". Damage, State & Date are the same but they do have different County names (Mercer & Henry). The remarks in the 'REMARKS' column are almost identical except for the county names. The 'REFNUM' column refers to 808182 and 808183 consecutively. However, the two counties are close to each other so it's possible each sustained similar damage in this event.

subset(data96, PROPDMG == 5000 & EVTYPE == 'FLASH FLOOD')

## STATE\_\_ BGN\_DATE BGN\_TIME TIME\_ZONE COUNTY COUNTYNAME STATE  
## 808182 17 2010-05-13 03:00:00 AM CST 131 MERCER IL  
## 808183 17 2010-05-13 03:05:00 AM CST 73 HENRY IL  
## EVTYPE BGN\_RANGE BGN\_AZI BGN\_LOCATI END\_DATE  
## 808182 FLASH FLOOD 8 WNW ELIZA 5/13/2010 0:00:00  
## 808183 FLASH FLOOD 3 WSW GREEN ROCK 5/13/2010 0:00:00  
## END\_TIME COUNTY\_END COUNTYENDN END\_RANGE END\_AZI END\_LOCATI  
## 808182 07:00:00 AM 0 NA 3 NNE SWEDONA  
## 808183 07:05:00 AM 0 NA 2 NNW COLONA  
## LENGTH WIDTH F MAG FATALITIES INJURIES PROPDMG PROPDMGEXP CROPDMG  
## 808182 0 0 NA 0 0 0 5000 K 0  
## 808183 0 0 NA 0 0 0 5000 K 0  
## CROPDMGEXP WFO STATEOFFIC  
## 808182 K DVN ILLINOIS, Northwest  
## 808183 K DVN ILLINOIS, Northwest  
## ZONENAMES  
## 808182   
## 808183   
## LATITUDE LONGITUDE LATITUDE\_E LONGITUDE\_  
## 808182 4119 9106 4119 9025  
## 808183 4127 9025 4130 9021  
## REMARKS  
## 808182 EPISODE NARRATIVE: A warm front slowly lifted northward from central Missouri into eastern Iowa and northern Illinois May 12th and 13th. Several rounds of showers and thunderstorms moved across the area producing widespread torrential rains and some hail. As a result of the heavy rains, flash flooding was common. There were two corridors of heavy rain; one across Highway 20 with amounts around 2 to 3 inches and a second heavier corridor of 2 to 6 inches from Memphis, Missouri to near Geneseo, Illinois. Temperatures in the upper 40s to middle 60s on May 12 continued to rise during the early morning hours May 13, reaching into the lower 50s to upper 60s by daybreak.EVENT NARRATIVE: Heavy rains resulted in widespread flash flooding of roads and highways across all of Mercer County, especially the eastern half during the morning of May 13.  
## 808183 EPISODE NARRATIVE: A warm front slowly lifted northward from central Missouri into eastern Iowa and northern Illinois May 12th and 13th. Several rounds of showers and thunderstorms moved across the area producing widespread torrential rains and some hail. As a result of the heavy rains, flash flooding was common. There were two corridors of heavy rain; one across Highway 20 with amounts around 2 to 3 inches and a second heavier corridor of 2 to 6 inches from Memphis, Missouri to near Geneseo, Illinois. Temperatures in the upper 40s to middle 60s on May 12 continued to rise during the early morning hours May 13, reaching into the lower 50s to upper 60s by daybreak.EVENT NARRATIVE: Heavy rains resulted in widespread flash flooding of roads and highways across all of Henry County, especially the northwest half during the morning of May 13.  
## REFNUM  
## 808182 808182  
## 808183 808183

Similarly, there's a double entry for flood damage in Napa Dec 2005/ Jan 2006. One entry shows damage of 115M and a second entry has for 115B damage (REFNUM 567221 & 605943). A quick query on-line indicates that a more accurate value for this entry is 115 Million. As a consequence we will remove 'REFNUM' = 605943.

subset(data96, REFNUM == '567221' | REFNUM == '605943')

## STATE\_\_ BGN\_DATE BGN\_TIME TIME\_ZONE COUNTY COUNTYNAME STATE  
## 567251 6 2005-12-31 07:05:00 PM PST 55 NAPA CA  
## 605953 6 2006-01-01 12:00:00 AM PST 55 NAPA CA  
## EVTYPE BGN\_RANGE BGN\_AZI BGN\_LOCATI END\_DATE END\_TIME  
## 567251 FLOOD 0 COUNTYWIDE 12/31/2005 0:00:00 11:59:00 PM  
## 605953 FLOOD 0 COUNTYWIDE 1/1/2006 0:00:00 07:00:00 AM  
## COUNTY\_END COUNTYENDN END\_RANGE END\_AZI END\_LOCATI LENGTH WIDTH F  
## 567251 0 NA 0 COUNTYWIDE 0 0 NA  
## 605953 0 NA 0 COUNTYWIDE 0 0 NA  
## MAG FATALITIES INJURIES PROPDMG PROPDMGEXP CROPDMG CROPDMGEXP WFO  
## 567251 0 0 0 115 M 32.5 M MTR  
## 605953 0 0 0 115 B 32.5 M MTR  
## STATEOFFIC ZONENAMES LATITUDE LONGITUDE LATITUDE\_E  
## 567251 CALIFORNIA, Western 3828 12218 3828  
## 605953 CALIFORNIA, Western 3828 12218 3828  
## LONGITUDE\_  
## 567251 12218  
## 605953 12218  
## REMARKS  
## 567251 Severe Flooding occurred as the Napa River exceeded flood stage at both St.Helena and in the City of Napa. The Napa Creek in downtown Napa also flooded, damaging the entire business district. City and Parks Department in Napa was hit with $6 million in damage alone. The City of Napa had 600 homes with moderate damage, 150 damaged businesses with costs of at least $70 million. More than 5 inches of rain fell on Napa in less than 24 hours. The flooding continued into the first few days of January 2006.  
## 605953 Major flooding continued into the early hours of January 1st, before the Napa River finally fell below flood stage and the water receeded. Flooding was severe in Downtown Napa from the Napa Creek and the City and Parks Department was hit with $6 million in damage alone. The City of Napa had 600 homes with moderate damage, 150 damaged businesses with costs of at least $70 million.  
## REFNUM  
## 567251 567221  
## 605953 605943

#### Event Type and Crop Damage

For property damage we're interested in the 'CROPDMG' column. Here is a summary of crop damage. And similar to property, there is also a multiplier column 'CROPDMGEXP' which may or may not increase these values.

summary(data96$CROPDMG) #Max = 5000

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.000 0.000 0.000 1.839 0.000 990.000

any(is.na(data96$CROPDMG))

## [1] FALSE

Again taking into account the multipliers as we did in the property damage data, we have K, M, B and entries with no multiplier. And like the PROPDMG there are no CROPDMG entries where there's no multiplier so we don't have to worry about missing some costs when we filter for valid multipliers later.

unique(data96$CROPDMGEXP)

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

#Check for entries that have significant PROPDMG values but no Multiplier   
a <- subset(data96, CROPDMGEXP == '')  
summary(a$CROPDMG)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0 0 0 0 0 0

#### Dealing with multipliers

For this report we are only concerned with valid entries with letters K, M, and B entered in the PROPDMGEXP & CROPDMGEXP columns.

Section 2.7 page 12 of the [NATIONAL WEATHER SERVICE INSTRUCTION](https://d396qusza40orc.cloudfront.net/repdata/peer2_doc/pd01016005curr.pdf) published by NOAA states that for damage "Estimates should be rounded to three significant digits, followed by an alphabetical character signifying the magnitude of the number, i.e., 1.55B for $1,550,000,000. Alphabetical characters used to signify magnitude include "K" for thousands, "M" for millions, and "B" for billions."

Also an earlier [Memorandum](http://www.nws.noaa.gov/wsom/manual/archives/NF429405.HTML#4.2.1) refers to estimations and multipliers in 2.2.4 (casualities) 2.2.5 and 4.2.1 specifically for entries and damaged estimates.

Therefore, estimating multipliers other than those stated, may result in unreliable and most likely guessed estimates would may impact the reports findings. We will also filter out any entries with no multipliers as discussed earlier.

#### Processing datasets for property and crop damage

Select relevant columns for property and crop damage

#Select relevant data  
damageData <- select(data96, BGN\_DATE, EVTYPE, PROPDMG, PROPDMGEXP, CROPDMG, CROPDMGEXP, REFNUM)

Here we will process the property damage selecting relevant columns and removing the double entry as discussed above.

#Filter dataset for Property Damage only.  
propDMG <- select(damageData, EVTYPE, PROPDMG, PROPDMGEXP, REFNUM)  
propDMG <- filter(propDMG, REFNUM != '605943') #remove invalid double entry

Now subset dataset with valid multipliers only (K, M or B).

##Remove any data without K, M or B and calculate damage estimates  
kprop <- propDMG[(propDMG$PROPDMGEXP == 'K'),]  
Mprop <- propDMG[(propDMG$PROPDMGEXP == 'M'),]  
Bprop <- propDMG[(propDMG$PROPDMGEXP == 'B'),]  
  
dataDMG <- bind\_rows(kprop, Mprop)  
dataDMG <- bind\_rows(dataDMG, Bprop)

Adding multiplier values to specific multipler letters.

dataDMG$EXP[dataDMG$PROPDMGEXP == 'K'] <- 1000  
dataDMG$EXP[dataDMG$PROPDMGEXP == 'M'] <- 1e+06  
dataDMG$EXP[dataDMG$PROPDMGEXP == 'B'] <- 1e+09

Process the crop damage selecting relevant columns.

#Filter datset for crop damage only  
cropDMG <- select(damageData, EVTYPE, CROPDMG, CROPDMGEXP)

Subset the dataset with valid multipliers only (K, M or B).

##Remove any data without K, M or B and calculate damage estimates  
kcrop <- cropDMG[(cropDMG$CROPDMGEXP == 'K'),]  
Mcrop <- cropDMG[(cropDMG$CROPDMGEXP == 'M'),]  
Bcrop <- cropDMG[(cropDMG$CROPDMGEXP == 'B'),]  
  
dataCropDMG <- bind\_rows(kcrop, Mcrop)  
dataCropDMG <- bind\_rows(dataCropDMG, Bcrop)

Add the multiplier value to specific multipler letters.

#Caculate new values using multiplers  
dataCropDMG$EXP[dataCropDMG$CROPDMGEXP == 'K'] <- 1000  
dataCropDMG$EXP[dataCropDMG$CROPDMGEXP == 'M'] <- 1e+06  
dataCropDMG$EXP[dataCropDMG$CROPDMGEXP == 'B'] <- 1e+09

### Results

##### The top weather events most harmful with respect to population health aross the United States

Estimate and identify events with the most fatalities.

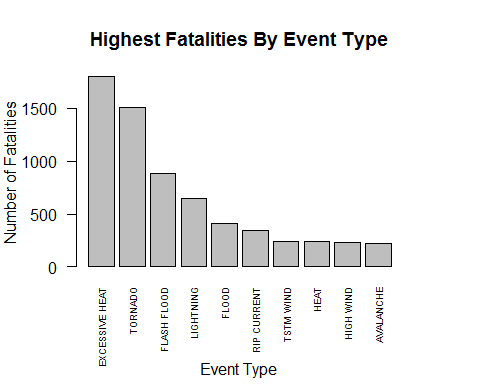
#Identify Fatalities caused by Events   
sumFatalities <- group\_by(stormData, EVTYPE) %>%  
 summarise(sum=sum(FATALITIES)) %>%  
 arrange(desc(sum)) %>%  
 print

## Source: local data frame [516 x 2]  
##   
## EVTYPE sum  
## 1 EXCESSIVE HEAT 1797  
## 2 TORNADO 1511  
## 3 FLASH FLOOD 887  
## 4 LIGHTNING 651  
## 5 FLOOD 414  
## 6 RIP CURRENT 340  
## 7 TSTM WIND 241  
## 8 HEAT 237  
## 9 HIGH WIND 235  
## 10 AVALANCHE 223  
## .. ... ...

sumFatalities <- head(sumFatalities,10) #top ten events

Create a Plot for Fatalities.

sumFatalities$EVTYPE <- as.character(sumFatalities$EVTYPE)  
  
par(mar=c(6,4,4,4))  
barplot(sumFatalities$sum, names.arg=sumFatalities$EVTYPE, cex.names=0.55,font.lab=1.8,las=2,main='Highest Fatalities By Event Type',  
 ylab='Number of Fatalities', xlab='', cex.lab = 1)  
mtext(side = 1, text = "Event Type", line = 4.7)

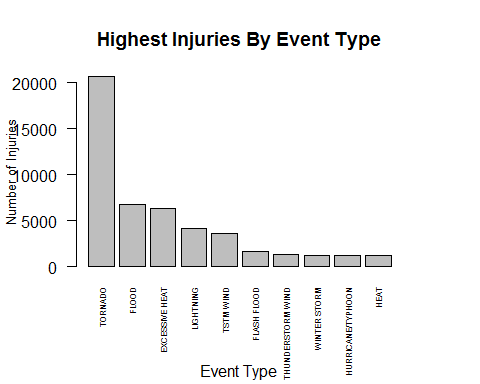
  
 Estimate and identify events with most injuries.

#Identify Injuries caused by Events  
sumInjuries <- group\_by(stormData, EVTYPE) %>%  
 summarise(sum=sum(INJURIES)) %>%  
 arrange(desc(sum)) %>%  
 print

## Source: local data frame [516 x 2]  
##   
## EVTYPE sum  
## 1 TORNADO 20667  
## 2 FLOOD 6758  
## 3 EXCESSIVE HEAT 6391  
## 4 LIGHTNING 4141  
## 5 TSTM WIND 3629  
## 6 FLASH FLOOD 1674  
## 7 THUNDERSTORM WIND 1400  
## 8 WINTER STORM 1292  
## 9 HURRICANE/TYPHOON 1275  
## 10 HEAT 1222  
## .. ... ...

Plot Injuries caused by Weather Events.

#Identify top 10 Events  
sumInjuries <- head(sumInjuries,10)  
sumInjuries$EVTYPE <- as.character(sumInjuries$EVTYPE)  
  
par(mar=c(6,4,4,4))  
barplot(sumInjuries$sum, names.arg=sumInjuries$EVTYPE, cex.names=0.53, font.lab =1.9, las=2,main='Highest Injuries By Event Type',  
 ylab='Number of Injuries', xlab='', cex.lab=0.8)  
mtext(side=1, text= 'Event Type', line= 4.8)

  
 Estimate and identify events with most property damage.

topPropDmg <- mutate(dataDMG, Est = PROPDMG \* EXP) %>%  
 group\_by(EVTYPE) %>%  
 summarise(EstDamage = sum(Est)) %>%  
 arrange(desc(EstDamage)) %>%  
 print

## Source: local data frame [178 x 2]  
##   
## EVTYPE EstDamage  
## 1 HURRICANE/TYPHOON 69305840000  
## 2 STORM SURGE 43193536000  
## 3 FLOOD 28944833550  
## 4 TORNADO 24616945710  
## 5 FLASH FLOOD 15222203910  
## 6 HAIL 14595143420  
## 7 HURRICANE 11812819010  
## 8 TROPICAL STORM 7642475550  
## 9 HIGH WIND 5247860360  
## 10 WILDFIRE 4758667000  
## .. ... ...

Estimate and identify events with the most crop damage.

#Establish Events with most Damage  
topCropDmg <- mutate(dataCropDMG, Est = CROPDMG \* EXP) %>%  
 group\_by(EVTYPE) %>%  
 summarise(EstDamage = sum(Est)) %>%  
 arrange(desc(EstDamage)) %>%  
 print

## Source: local data frame [83 x 2]  
##   
## EVTYPE EstDamage  
## 1 DROUGHT 13367566000  
## 2 FLOOD 4974778400  
## 3 HURRICANE 2741410000  
## 4 HURRICANE/TYPHOON 2607872800  
## 5 HAIL 2476029450  
## 6 FLASH FLOOD 1334901700  
## 7 EXTREME COLD 1288973000  
## 8 FROST/FREEZE 1094086000  
## 9 HEAVY RAIN 728169800  
## 10 TROPICAL STORM 677711000  
## .. ... ...

To calculate overall economic damage we will merge crop and property damage.

mergeData <- merge(topPropDmg, topCropDmg, by.x='EVTYPE', by.y='EVTYPE',all=TRUE)  
mergeData[is.na(mergeData)] <- 0 # change NAs to zero.  
economicDmg <- rename(mergeData, 'PropDamage'=EstDamage.x, 'CropDamage'=EstDamage.y)  
economicDmg <- mutate(economicDmg, TotalDMG = PropDamage + CropDamage)  
  
economicDmg <- arrange(economicDmg, desc(TotalDMG))  
totalEconomicDmg <- head(economicDmg,10)

Now plot the the top weather events causing the most economic damage.

x <- rbind(totalEconomicDmg$PropDamage, totalEconomicDmg$CropDamage)  
rownames(x) <- c("Property Damage", "Crop Damage")  
colnames(x) <- totalEconomicDmg$EVTYPE  
  
  
par(mar=c(6,5,3,3))  
barplot(x, legend=rownames(x), ylab='', xlab='', las=2, cex.names=0.53,font.lab=1.8,cex.lab=0.5, main='Economic Damage by Event Type')  
mtext(side=1, text= 'Event Type', line= 4.8)  
mtext(side=2, text= 'Total Damage in $US', line= 4.1)

