RR: Assessing Health and Economic Impact Of Weather Events

Olagunju Ayodeji

February 3, 2019

## Synopsis of Study Results

This study 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.

Wind events, which include tornadoes and hurricanes, are by the far the most harmful in aggregate, causing over 100,000 injuries and 90 deaths over the course of this study. Though less frequent, severe heat events have the highest incidence of deaths and injuries per event. This study finds that severe rain and wind events are by far the most costly in terms of dollars spent to replace property and crop damage.

## Questions this study considers

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

## Notes about the environment used

This study was done using the following tools, including OS and Programming language versions

The study was conducted on a 32-bit Windows 7 machine with 2 cores.

R language was R version 3.5.2 (2018-12-20)

For publishing to rpubs.com, I used RStudio version 1.1.463

The full project may be found on Github at https://github.com/djolas/05-reproducible-research-assignment-2

## Data Processing

There will be categorization of event type that may be subjective. Not to mention the data collection of the weather events will be categorized on the field manually and hence subjected to human input errors.

setwd ("C:/Users/djolas/My Documents/data")

url <- "https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2"  
destfile <- "repdata%2Fdata%2FStormData.csv.bz2"  
download.file(url,destfile,mode="wb")  
url2 <- "https://d396qusza40orc.cloudfront.net/repdata%2Fpeer2\_doc%2Fpd01016005curr.pdf"  
destfile2 <- "repdata%2Fpeer2\_doc%2Fpd01016005curr.pdf"  
download.file(url2,destfile2,mode="wb")

StormData<- read.csv(bzfile(destfile), stringsAsFactors = FALSE)  
names(StormData)

## [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"

summary(StormData$FATALITIES)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.0000 0.0000 0.0000 0.0168 0.0000 583.0000

summary(StormData$INJURIES)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.0000 0.0000 0.0000 0.1557 0.0000 1700.0000

It seems that there are 985 unique types of events, I will endeavor to decrease this list a lot more I will also create a list for Question 1 for most harmful and Question 2 for damages PROPDMGEXP,CROPDMGEXP Also, I will only need certain columns FATALITIES INJURIES PROPDMG CROPDMG PROPDMGEXP CROPDMGEXP

KeepCol <- c("EVTYPE","FATALITIES","INJURIES","PROPDMG","CROPDMG","PROPDMGEXP","CROPDMGEXP")  
StormDataKeep <- subset(StormData, select = KeepCol)  
names(StormDataKeep)

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

StormDataKeepNZ <- subset(StormDataKeep, FATALITIES > 0 | INJURIES > 0 | PROPDMG > 0 | CROPDMG > 0)

StormDataKeepNZ$EVENT <- StormDataKeepNZ$EVTYPE  
StormDataKeepNZ$EVENT[grep("flood",StormDataKeepNZ$EVENT, ignore.case = T)] <-"FLOOD"  
StormDataKeepNZ$EVENT[grep("warm",StormDataKeepNZ$EVENT, ignore.case = T)] <-"HEAT"  
StormDataKeepNZ$EVENT[grep("freeze",StormDataKeepNZ$EVENT, ignore.case = T)] <-"COLD"  
StormDataKeepNZ$EVENT[grep("avala",StormDataKeepNZ$EVENT, ignore.case = T)] <-"AVALANCHE"  
StormDataKeepNZ$EVENT[grep("ice",StormDataKeepNZ$EVENT, ignore.case = T)] <-"COLD"  
StormDataKeepNZ$EVENT[grep("bitter",StormDataKeepNZ$EVENT, ignore.case = T)] <-"COLD"  
StormDataKeepNZ$EVENT[grep("BELOW NORMAL PRECIPITATION",StormDataKeepNZ$EVENT, ignore.case = T)] <-"DROUGHT"  
StormDataKeepNZ$EVENT[grep("BLIZZARD",StormDataKeepNZ$EVENT, ignore.case = T)] <-"SNOW"  
StormDataKeepNZ$EVENT[grep("SNOW",StormDataKeepNZ$EVENT, ignore.case = T)] <-"SNOW"  
StormDataKeepNZ$EVENT[grep("dry",StormDataKeepNZ$EVENT, ignore.case = T)] <-"DROUGHT"  
StormDataKeepNZ$EVENT[grep("BEACH",StormDataKeepNZ$EVENT, ignore.case = T)] <-"BEACH EROSION"  
StormDataKeepNZ$EVENT[grep("FIRE",StormDataKeepNZ$EVENT, ignore.case = T)] <-"FIRE"  
StormDataKeepNZ$EVENT[grep("COLD",StormDataKeepNZ$EVENT, ignore.case = T)] <-"COLD"  
StormDataKeepNZ$EVENT[grep("RAIN",StormDataKeepNZ$EVENT, ignore.case = T)] <-"RAIN"  
StormDataKeepNZ$EVENT[grep("CHILL",StormDataKeepNZ$EVENT, ignore.case = T)] <-"COLD"  
StormDataKeepNZ$EVENT[grep("wet",StormDataKeepNZ$EVENT, ignore.case = T)] <-"FLOOD"  
StormDataKeepNZ$EVENT[grep("frost",StormDataKeepNZ$EVENT, ignore.case = T)] <-"COLD"  
StormDataKeepNZ$EVENT[grep("freez",StormDataKeepNZ$EVENT, ignore.case = T)] <-"COLD"  
StormDataKeepNZ$EVENT[grep("gust",StormDataKeepNZ$EVENT, ignore.case = T)] <-"WIND"  
StormDataKeepNZ$EVENT[grep("hail",StormDataKeepNZ$EVENT, ignore.case = T)] <-"HAIL"  
StormDataKeepNZ$EVENT[grep("heat",StormDataKeepNZ$EVENT, ignore.case = T)] <-"HEAT"  
StormDataKeepNZ$EVENT[grep("hurricane",StormDataKeepNZ$EVENT, ignore.case = T)] <-"HURRICANE"  
StormDataKeepNZ$EVENT[grep("typhoon",StormDataKeepNZ$EVENT, ignore.case = T)] <-"HURRICANE"  
StormDataKeepNZ$EVENT[grep("ice",StormDataKeepNZ$EVENT, ignore.case = T)] <-"COLD"  
StormDataKeepNZ$EVENT[grep("icy",StormDataKeepNZ$EVENT, ignore.case = T)] <-"COLD"  
StormDataKeepNZ$EVENT[grep("LANDSLIDE",StormDataKeepNZ$EVENT, ignore.case = T)] <-"LANDSLIDES"  
StormDataKeepNZ$EVENT[grep("mud",StormDataKeepNZ$EVENT, ignore.case = T)] <-"MUDSLIDES"  
StormDataKeepNZ$EVENT[grep("high temp",StormDataKeepNZ$EVENT, ignore.case = T)] <-"HEAT"  
StormDataKeepNZ$EVENT[grep("TROPICAL STORM",StormDataKeepNZ$EVENT, ignore.case = T)] <-"TROPICAL STORM"  
StormDataKeepNZ$EVENT[grep("light",StormDataKeepNZ$EVENT, ignore.case = T)] <-"THUNDERSTORM"  
StormDataKeepNZ$EVENT[grep("tstm",StormDataKeepNZ$EVENT, ignore.case = T)] <-"THUNDERSTORM"  
StormDataKeepNZ$EVENT[grep("torn",StormDataKeepNZ$EVENT, ignore.case = T)] <-"TORNADO"  
StormDataKeepNZ$EVENT[grep("tide",StormDataKeepNZ$EVENT, ignore.case = T)] <-"FLOOD"  
StormDataKeepNZ$EVENT[grep("tsu",StormDataKeepNZ$EVENT, ignore.case = T)] <-"FLOOD"  
StormDataKeepNZ$EVENT[grep("thun",StormDataKeepNZ$EVENT, ignore.case = T)] <-"THUNDERSTORM"  
StormDataKeepNZ$EVENT[grep("tsu",StormDataKeepNZ$EVENT, ignore.case = T)] <-"FLOOD"  
StormDataKeepNZ$PropertyDamageAmt <- StormDataKeepNZ$PROPDMG  
MultLookup <- c(M = 10^6, m = 10^6, K = 10^3, k = 10^3, B = 10^9, b = 10^9)  
StormDataKeepNZ$PropertyDamageAmt <- StormDataKeepNZ$PROPDMG \* MultLookup [as.character(StormDataKeepNZ$PROPDMGEXP)]  
StormDataKeepNZ$CropDamageAmt <- StormDataKeepNZ$CROPDMG \* MultLookup [as.character(StormDataKeepNZ$CROPDMGEXP)]  
StormDataKeepNZ$TotDamageAmt <- StormDataKeepNZ$PropertyDamageAmt + StormDataKeepNZ$CropDamageAmt

====================================================================

# RESULTS

Now we will aggregate the results and output the top 10 events graphically. This will answer the 2 questions

library(plyr)  
SFatal <- ddply(StormDataKeepNZ, .(EVENT), summarize, DEATHS = sum(FATALITIES, na.rm = TRUE))  
SInjury <- ddply(StormDataKeepNZ, .(EVENT), summarize, INJURED = sum(INJURIES, na.rm = TRUE))  
SPropDam <- ddply(StormDataKeepNZ, .(EVENT), summarize, PROPERTYCOST = sum(PropertyDamageAmt, na.rm = TRUE))  
SCropDam <- ddply(StormDataKeepNZ, .(EVENT), summarize, CROPCOST = sum(CropDamageAmt, na.rm = TRUE))  
STotDam <- ddply(StormDataKeepNZ, .(EVENT), summarize, TOTALCOST = sum(TotDamageAmt, na.rm = TRUE))  
SDEATH10 <- head(SFatal[order(-SFatal$DEATHS),],10)  
SINJURED10 <- head(SInjury [order(-SInjury$INJURED),],10)  
SPropDam10 <- head(SPropDam[order(-SPropDam$PROPERTYCOST),],10)  
SCropDam10 <- head(SCropDam[order(-SCropDam$CROPCOST),],10)  
STotDam10 <- head(STotDam[order(-STotDam$TOTALCOST),],10)  
SDEATH10

## EVENT DEATHS  
## 106 TORNADO 5636  
## 35 HEAT 3178  
## 25 FLOOD 1569  
## 105 THUNDERSTORM 1542  
## 12 COLD 566  
## 87 RIP CURRENT 368  
## 97 SNOW 264  
## 53 HIGH WIND 248  
## 4 AVALANCHE 225  
## 126 WINTER STORM 206

SINJURED10

## EVENT INJURED  
## 106 TORNADO 91407  
## 105 THUNDERSTORM 14679  
## 35 HEAT 9243  
## 25 FLOOD 8738  
## 12 COLD 2538  
## 97 SNOW 1958  
## 24 FIRE 1608  
## 33 HAIL 1467  
## 61 HURRICANE 1333  
## 126 WINTER STORM 1321

SPropDam10

## EVENT PROPERTYCOST  
## 25 FLOOD 172324770320  
## 61 HURRICANE 85356410010  
## 106 TORNADO 56993098180  
## 99 STORM SURGE 43323536000  
## 33 HAIL 17619970720  
## 105 THUNDERSTORM 11859474310  
## 24 FIRE 8501628500  
## 108 TROPICAL STORM 7714390550  
## 126 WINTER STORM 6688497250  
## 53 HIGH WIND 5270046260

SCropDam10

## EVENT CROPCOST  
## 17 DROUGHT 13972581000  
## 25 FLOOD 12527979100  
## 12 COLD 8452940850  
## 61 HURRICANE 5516117800  
## 33 HAIL 3114212850  
## 105 THUNDERSTORM 1218945740  
## 35 HEAT 904479280  
## 85 RAIN 806162800  
## 108 TROPICAL STORM 694896000  
## 53 HIGH WIND 638571300

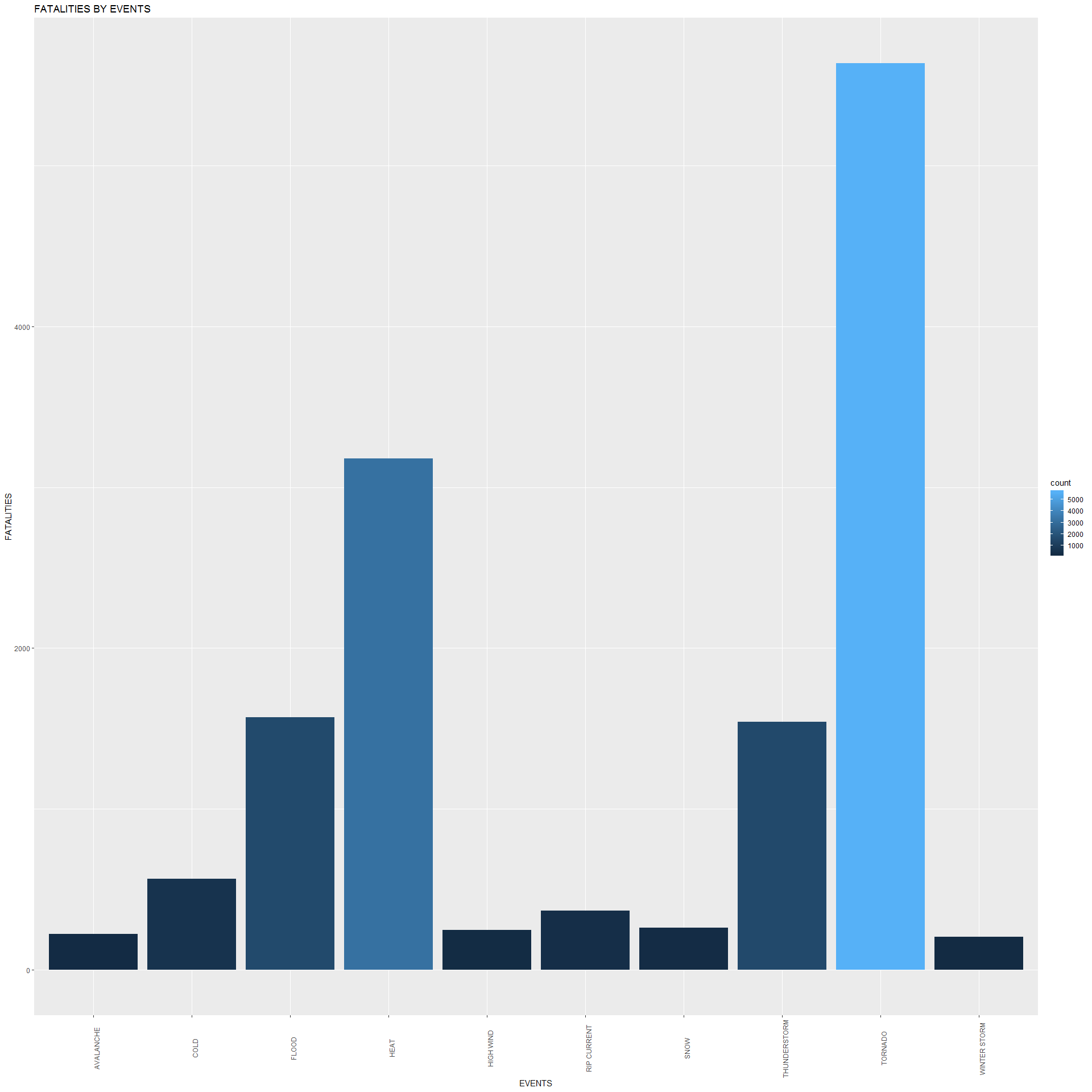
STotDam10

## EVENT TOTALCOST  
## 25 FLOOD 162549350740  
## 61 HURRICANE 44330000800  
## 106 TORNADO 16520165550  
## 33 HAIL 11653045140  
## 12 COLD 7002475700  
## 105 THUNDERSTORM 5805209380  
## 24 FIRE 3838549570  
## 53 HIGH WIND 3057666640  
## 17 DROUGHT 1886540000  
## 108 TROPICAL STORM 1530352350

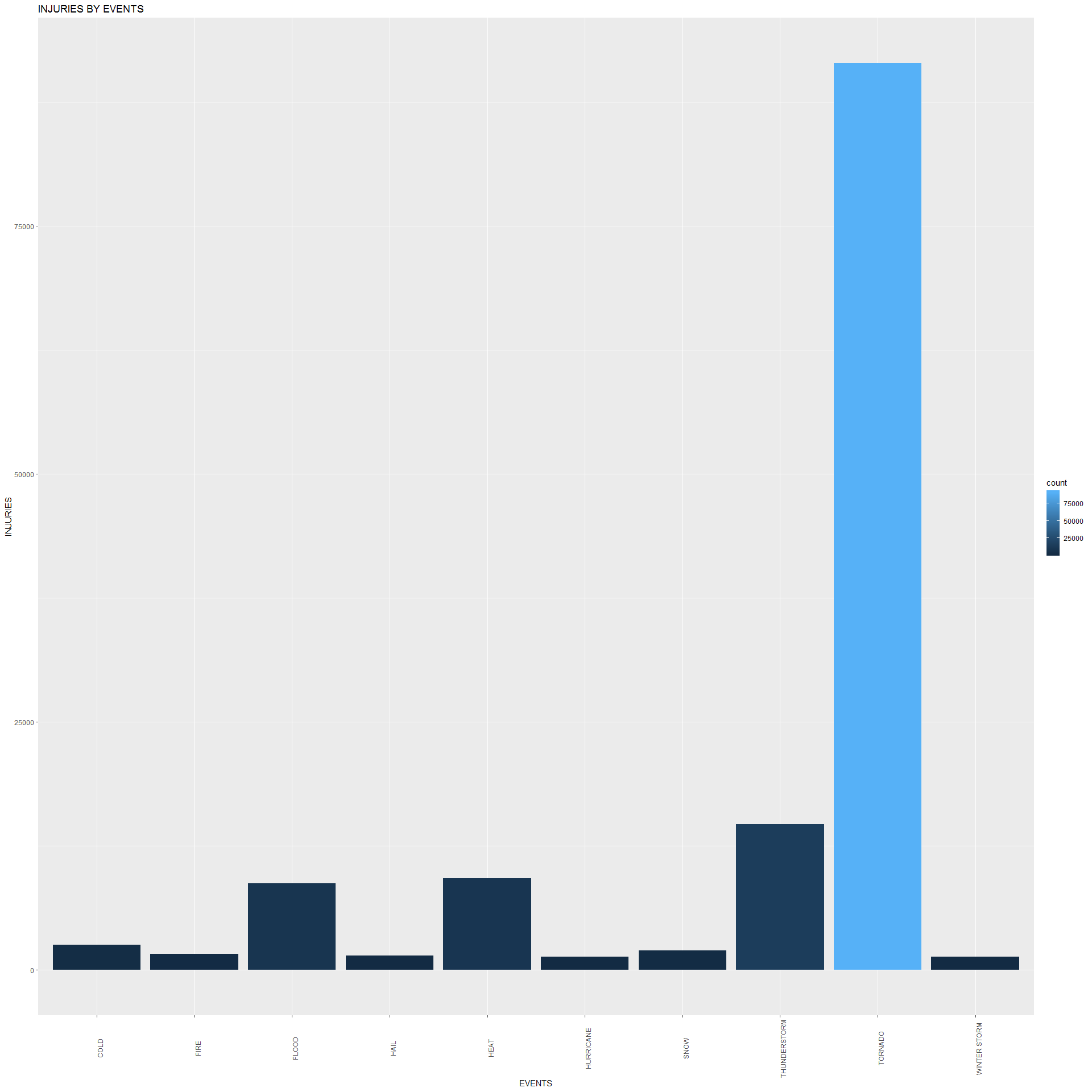
=========================================================

## Graphing the results

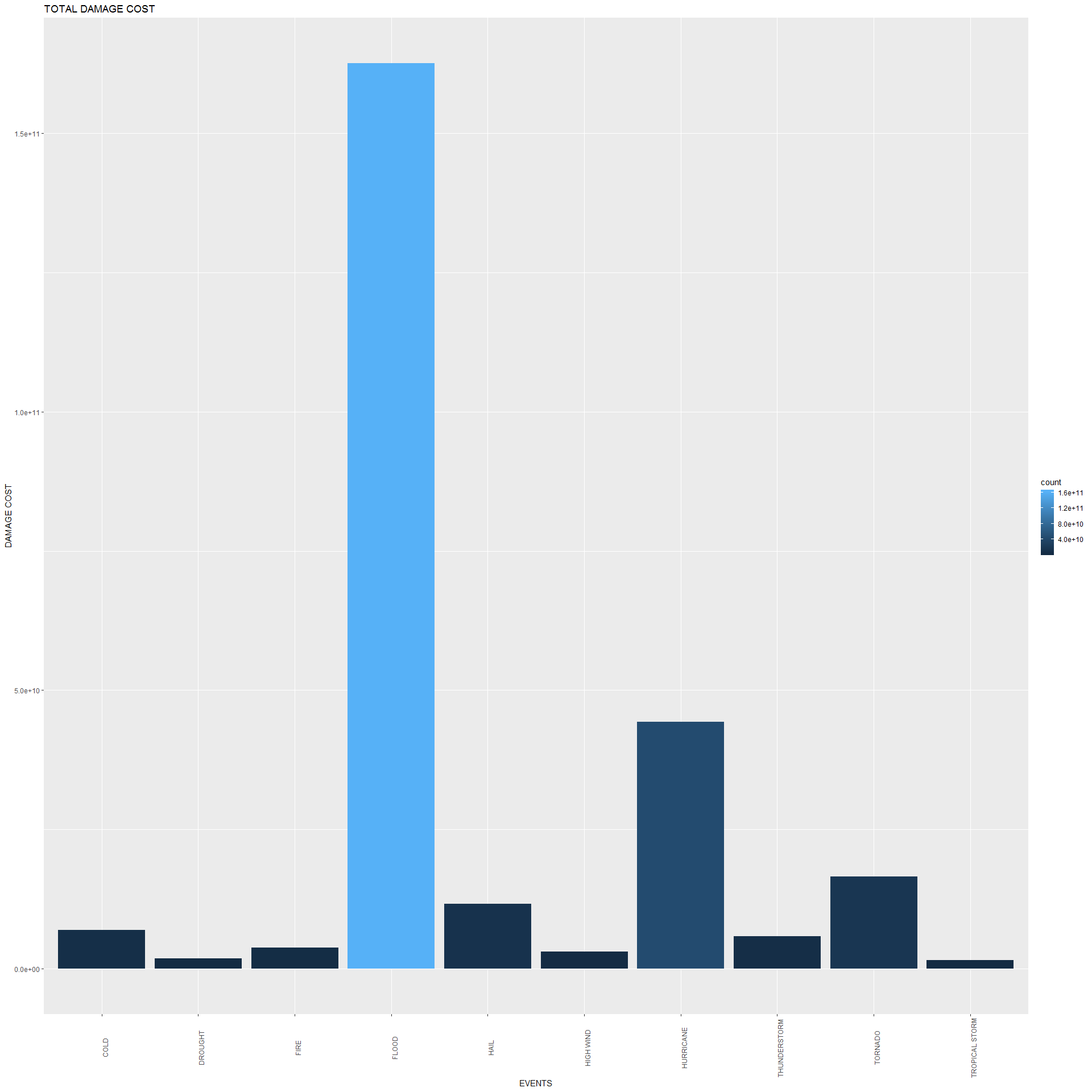
library (ggplot2)  
ggplot(SDEATH10,aes(EVENT)) +   
 geom\_bar(aes(weight=DEATHS, fill =..count..)) +   
 xlab("EVENTS") + theme(axis.text.x=element\_text(angle = 90)) +   
 ylab("FATALITIES") +   
 ggtitle("FATALITIES BY EVENTS")



ggplot (SINJURED10,aes(EVENT)) +   
 geom\_bar(aes(weight=INJURED, fill =..count..),) +   
 xlab("EVENTS") + theme(axis.text.x=element\_text(angle = 90)) +   
 ylab("INJURIES") +   
 ggtitle("INJURIES BY EVENTS")



ggplot (STotDam10,aes(EVENT)) +   
 geom\_bar(aes(weight=TOTALCOST, fill =..count..)) +   
 xlab("EVENTS") + theme(axis.text.x=element\_text(angle = 90)) +   
 ylab("DAMAGE COST") +   
 ggtitle("TOTAL DAMAGE COST")



================================================================================

## SUMMARY

The main causes of property damage are floods, hurricanes and tornadoes Crop damages are generally caused by drought, flood, and cold

Fatalities are far and away caused by tornadoes and thunderstorms Injuries are also far and away caused by tornadoes, with heat, thunderstorms.