Reproducible Research Final Project

Nicolas Gonzalez Rojas

23/10/2020

Overview

The goal of the assignment is to explore the NOAA Storm Database and explore the effects of severe weather events on both population and economy.The database covers the time period between 1950 and November 2011.

The following analysis investigates which types of severe weather events are most harmful on:

1. Health (injuries and fatalities)
2. Property and crops (economic consequences)

Data Processing

Data loading

Download the raw data file and extract the data into a dataframe.Then convert to a data.table

```
link <- "https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2"
download.file(link, destfile = "dataset.csv.bz2")
StormDB <- read.csv("dataset.csv.bz2")
library(ggplot2)
```

Examining Column Names

```
colnames(StormDB)
```

```
##  [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"
```

Identify Event Type Labels that should be scrubbed.

```
event_types <- as.data.frame(table(StormDB$EVTYPE))
event_types <- event_types[order(event_types$Var1), ]
```

Clean up a majority of Identified Names In order to properly count and categorize records that have possible multiple events, records that possess an ampersand, slash, or 'and' will be labeled as a multiple event.

The naming of the event is to be done on the general overriding idea behind the event. For example, wind 65+ will be categorized the same as wind 45+ because both specific events deal with the event type of wind. This is done over several different instances.

```
StormDB$EVTYPE <- as.character(StormDB$EVTYPE)
StormDB$EVTYPE[grepl("/|&|and", StormDB$EVTYPE,ignore.case = TRUE)] <- "Multiple Event"
StormDB$EVTYPE[grepl("volc", StormDB$EVTYPE,ignore.case = TRUE)] <- "Volcano"
StormDB$EVTYPE[grepl("wind|wnd", StormDB$EVTYPE,ignore.case = TRUE)] <- "WIND"
StormDB$EVTYPE[grepl("funnel|tornado", StormDB$EVTYPE,ignore.case = TRUE)] <- "Tornado"
StormDB$EVTYPE[grepl("glaze", StormDB$EVTYPE,ignore.case = TRUE)] <- "Glaze"
StormDB$EVTYPE[grepl("hail", StormDB$EVTYPE,ignore.case = TRUE)] <- "Hail"
StormDB$EVTYPE[grepl("dust", StormDB$EVTYPE,ignore.case = TRUE)]  <- "DUST"
StormDB$EVTYPE[grepl("flood", StormDB$EVTYPE,ignore.case = TRUE)] <- "FLOOD"
StormDB$EVTYPE[grepl("ic(e|y)", StormDB$EVTYPE,ignore.case = TRUE)] <- "Ice"
StormDB$EVTYPE[grepl("fire|smoke", StormDB$EVTYPE,ignore.case = TRUE)] <- "FIRE"
StormDB$EVTYPE[grepl("thunder", StormDB$EVTYPE,ignore.case = TRUE)] <- "Thunder Storm"
StormDB$EVTYPE[grepl("slide|eros", StormDB$EVTYPE,ignore.case = TRUE)] <- "Erosion"
StormDB$EVTYPE[grepl("rain", StormDB$EVTYPE,ignore.case = TRUE)] <- "Rain"
StormDB$EVTYPE[grepl("freez|cold|snow|chill|winter", StormDB$EVTYPE,ignore.case = TRUE)] <- "Cold Weath
StormDB$EVTYPE[grepl("TROPICAL.STORM", StormDB$EVTYPE,ignore.case = TRUE)] <- "TROPICAL STORM"
StormDB$EVTYPE[grepl("heat", StormDB$EVTYPE,ignore.case = TRUE)] <- "Heat"
StormDB$EVTYPE[grepl("(hurri|opal)", StormDB$EVTYPE,ignore.case = TRUE)] <- "Hurricane"
```

Seperate Data To Relevant Data for Question

```
health <- StormDB[,(c(8,23:24))]
property<-StormDB[,c(8,25:28)]
```

Property Data Processing

```
table(property$PROPDMGEXP)
```

```
##
##             -      ?      +      0      1      2      3      4      5      6
## 465934      1      8      5    216     25     13      4      4     28      4
##      7      8      B      h      H      K      m      M
##      5      1     40      1      6 424665      7  11330
```

```
table(property$CROPDMGEXP)
```

```
##
##             ?      0      2      B      k      K      m      M
## 618413      7     19      1      9     21 281832      1   1994
```

```
property$PROPDMGEXP<-factor(property$PROPDMGEXP,levels=c("H","K","M","B","h","m","O"))
property$PROPDMGEXP[is.na(property$PROPDMGEXP)] <- "O"

property$CROPDMGEXP<-factor(property$CROPDMGEXP,levels=c("K","M","B","k","m","O"))
property$CROPDMGEXP[is.na(property$CROPDMGEXP)] <- "O"
```

Convert the magnitude into the multiplier used for calculating damage amount.

Using the following key to identify the multiplier for the orders of magnitude.

1. o(one) = 1
2. h(undred)=100
3. k(thousand)=1000
4. m(million)=1000000
5. b(billion)=1000000000

```
property$PROPDMGEXP <- as.character(property$PROPDMGEXP)
property$CROPDMGEXP <- as.character(property$CROPDMGEXP)

property$PROPDMGMLT <- 0
property$CROPDMGMLT <- 0

property$PROPDMGMLT[grepl("h", property$PROPDMGEXP,ignore.case = TRUE)]<-100
property$PROPDMGMLT[grepl("k", property$PROPDMGEXP,ignore.case = TRUE)]<-1000
property$PROPDMGMLT[grepl("m", property$PROPDMGEXP,ignore.case = TRUE)]<-1000000
property$PROPDMGMLT[grepl("b", property$PROPDMGEXP,ignore.case = TRUE)]<-1000000000
property$PROPDMGMLT[grepl("o", property$PROPDMGEXP,ignore.case = TRUE)]<-1

property$CROPDMGMLT[grepl("k", property$CROPDMGEXP,ignore.case = TRUE)]<-1000
property$CROPDMGMLT[grepl("m", property$CROPDMGEXP,ignore.case = TRUE)]<-1000000
property$CROPDMGMLT[grepl("b", property$CROPDMGEXP,ignore.case = TRUE)]<-1000000000
property$CROPDMGMLT[grepl("o", property$CROPDMGEXP,ignore.case = TRUE)]<-1

property$PROPDMG <- property$PROPDMG * property$PROPDMGMLT
property$CROPDMG <- property$CROPDMG * property$CROPDMGMLT
property$total <- property$PROPDMG + property$CROPDMG
```

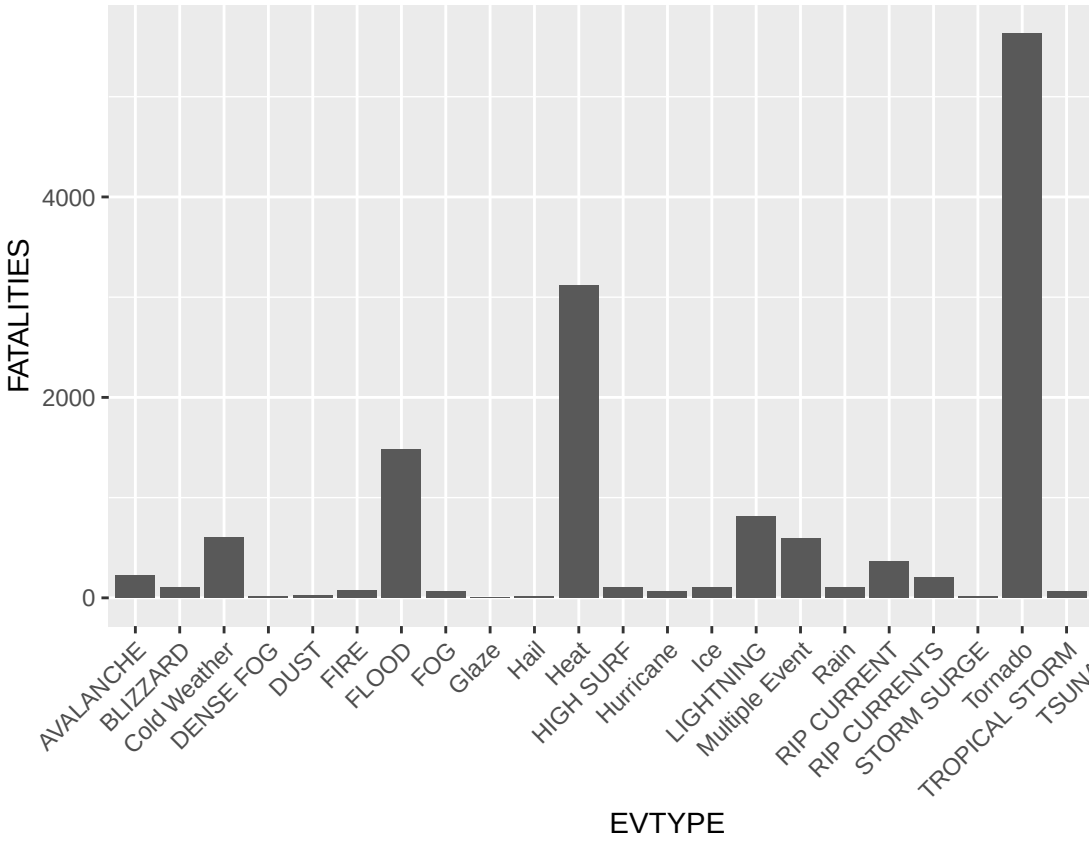
Results

Population Health Question

```
health.totals <- aggregate(cbind(FATALITIES,INJURIES) ~ EVTYPE, data = health, sum, na.rm=TRUE)
health.totals$TOTAL <- health.totals$FATALITIES + health.totals$INJURIES
health.totals <- health.totals[order(-health.totals$TOTAL), ]
health.totals <- health.totals[1:25,]
```

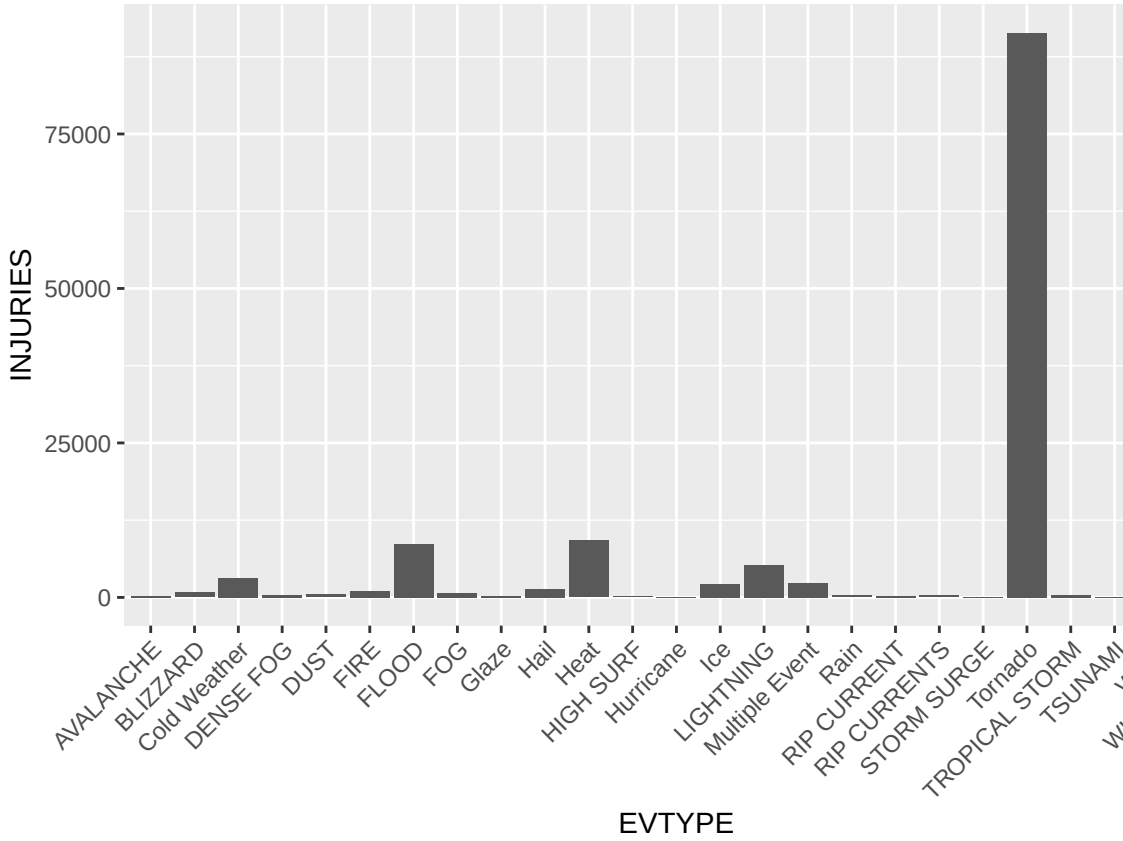
Health Totals

```
## FATALITIES
ggplot(health.totals, aes(EVTYPE, FATALITIES)) + geom_bar(stat = "identity") + theme(plot.title = eleme
```



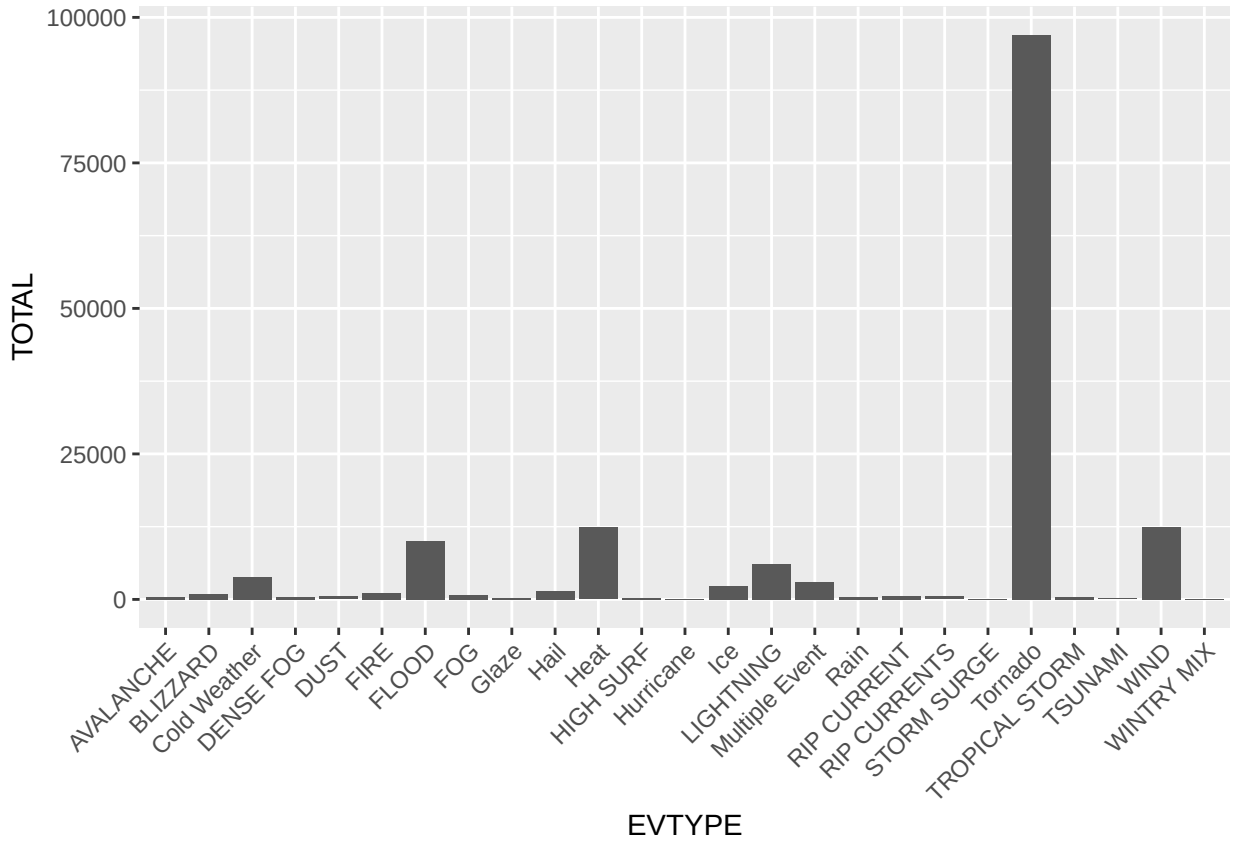
PLOT OF FATALITIES

```
## INJURIES
ggplot(health.totals, aes(EVTYPE, INJURIES )) + geom_bar(stat = "identity") + theme(plot.title = elemen
```



PLOT OF INJURIES

```
## TOTAL
ggplot(health.totals, aes(EVTYPE, TOTAL )) + geom_bar(stat = "identity") + theme(plot.title = element_t
```



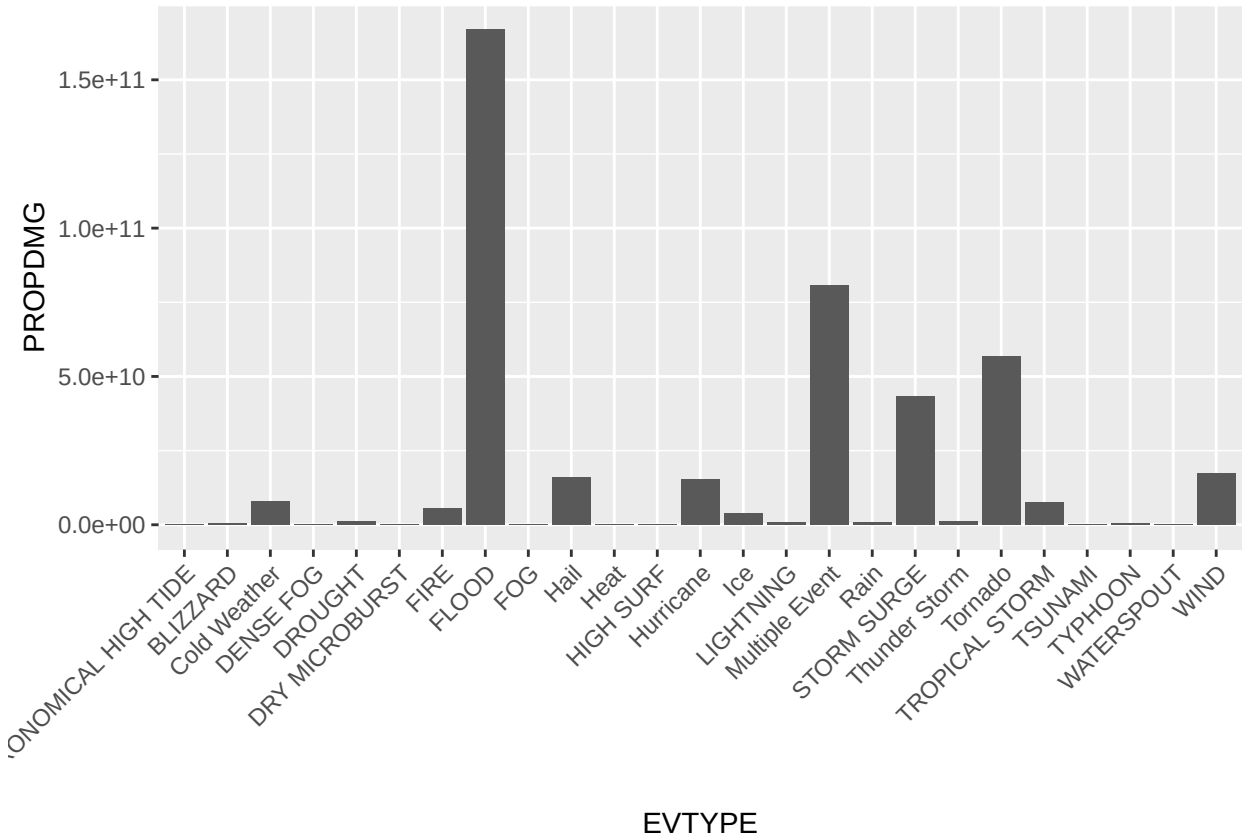
TOTAL

It is easily said that tornado's cause the largest weather-related risk to the overall population health. However, the averages of the events tell a different story about the most deadly single weather events. This will require additional research and analysis to properly identify which event has the worst outcomes for popluation health.

Economic Impact We will begin to look at the Economic Impact of certain types of events.

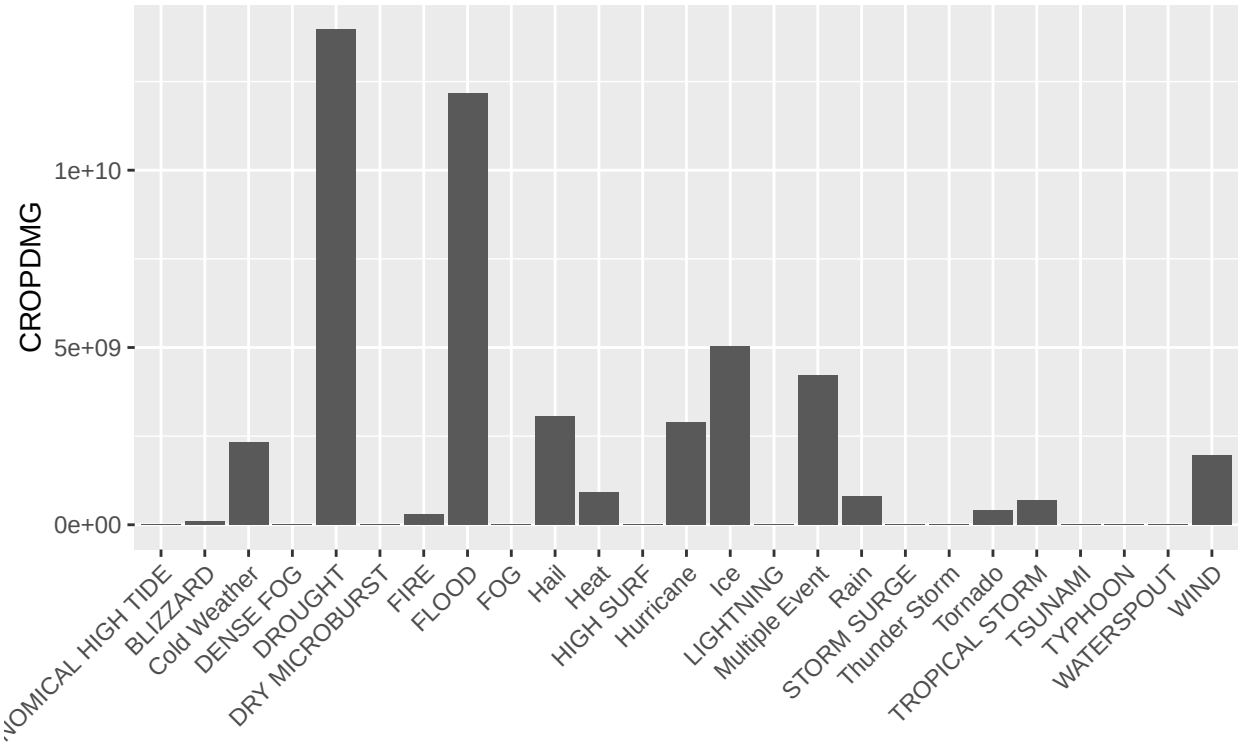
```
economic.total <- aggregate(cbind(PROPDMG,CROPDMG, total) ~ EVTYPE, data = property, sum, na.rm=TRUE)
economic.crop <- economic.total[order(-economic.total$CROPDMG), ]
economic.crop <- economic.crop[1:25,]

economic.prop <- economic.total[order(-economic.total$PROPDMG), ]
economic.prop <- economic.prop[1:25,]
```

```
ggplot(economic.prop, aes(EVTYPE, PROPDMG)) + geom_bar(stat = "identity") + theme(plot.title = element_
```



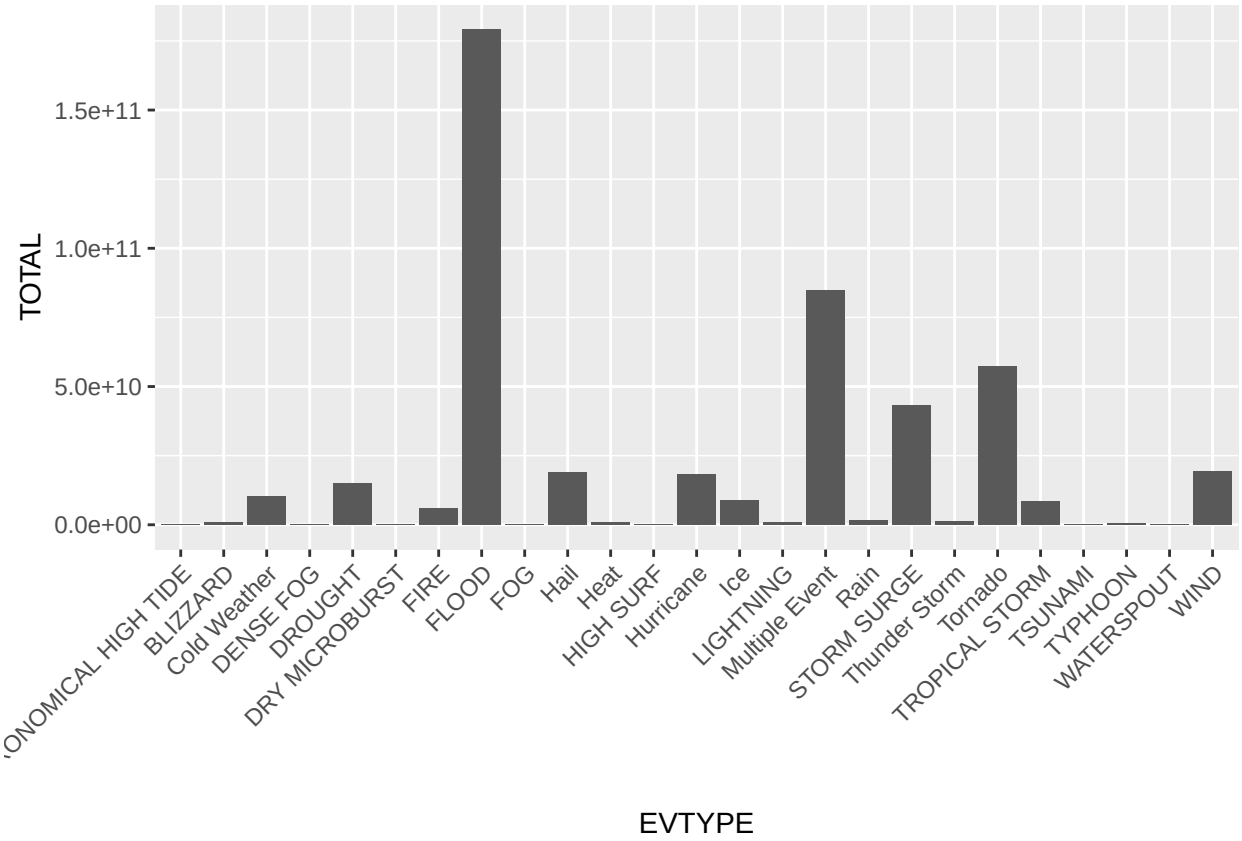
PROPCOST

```
ggplot(economic.prop, aes(EVTYPE, CROPDMG)) + geom_bar(stat = "identity") + theme(plot.title = element_
```



CROPCOST

```
## TOTAL
ggplot(economic.prop, aes(EVTYPE, total)) + geom_bar(stat = "identity") + theme(plot.title = element_te
```



TOTAL

While drought has the largest impact on crops, it is easy to see that flooding produces the largest overall weather-related impact to the economy. With the cost fully associated with crop destruction is not in the scope of this analysis, futher research is required to determine the full economic impact of one of these weather related events.