NOAA Storm Database Analysis

Michael Harrison

February 22, 2017

Introduction

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

This 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.

Your data analysis must address the following questions:

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

Data Processing

Cleaning

Data cleaning and organization will depend upon the following packages: data.table, dplyr, lubridate, and ggplot2(which is loaded later in the analysis to avoid masking issues with dplyr). Data from csv file read in using fread function of the data.table package for speed. All column names are set to lower case for ease in scripting; state columns renamed for clarity. Dates are transformed using the lubridate package function to set column to unambiguous date format. Data is then subset by date, starting on the first of January 1996, as NOAA began recording all event types.

```
library(data.table); library(dplyr);
library(lubridate);
options(scipen = 999)

stormdt <- fread("StormData.csv")

colnames(stormdt) <- tolower(colnames(stormdt))
colnames(stormdt)[1] <- "fips"
colnames(stormdt)[7] <- "state"

#change date formats
stormdt$bgn_date <- mdy_hms(stormdt$bgn_date)
stormdt$end_date <- mdy_hms(stormdt$end_date)
stormdt <- stormdt[bgn_date >= "1996-01-01"]
```

The following script takes the factor of property and crop damage and applies it to the damage standardize the value.

```
factordmg <- function(DT){
        DT[propdmgexp == "H" | propdmgexp == "h", propdmg := propdmg * 100]
        DT[propdmgexp == "K" | propdmgexp == "k", propdmg := propdmg * 1000]
```

```
        DT[propdmgexp == "M" | propdmgexp == "m", propdmg := propdmg * 1E6]
        DT[propdmgexp == "B" | propdmgexp == "b", propdmg := propdmg * 1E9]
        for(i in 1:9){stormdt[propdmgexp == i, propdmg := propdmg * 10^i]}

        DT[cropdmgexp == "H" | cropdmgexp == "h", cropdmg := cropdmg * 100]
        DT[cropdmgexp == "K" | cropdmgexp == "k", cropdmg := cropdmg * 1000]
        DT[cropdmgexp == "M" | cropdmgexp == "m", cropdmg := cropdmg * 1E6]
        DT[cropdmgexp == 0, cropdmg := cropdmg * 10^0]
        DT[cropdmgexp == 2, cropdmg := cropdmg * 10^2]
}

stormdt <- factordmg(stormdt)
```

Create new datetime columns for event begining and event ending.

```
stormdt$bgn_datetime <- as.POSIXct(paste(stormdt$bgn_date, stormdt$end_time),
                                   format = "%Y-%m-%d %H:%M:%S")
stormdt$end_datetime <- as.POSIXct(paste(stormdt$end_date, stormdt$end_time),
                                   format = "%Y-%m-%d %H:%M:%S")
```

The data is reshaped to eliminate column variables unneeded for the purposes of the analysis using the select function of dplyr. The data is then filtered to remove records where fatalities, injuries, property damage, or crop damage amounted to zero.

```
stormdt <- stormdt %>%
        select(fips, state, county, countyname, evtype,
               bgn_datetime, end_datetime,
               length, width, f, mag,
               fatalities, injuries, propdmg, cropdmg) %>%
        filter(fatalities > 0 | injuries > 0 | propdmg > 0 | cropdmg > 0)
```

Examining the unique values in the evtype variable returns nearly 1000 types of events, though has a set list of 48 event types for recording purposes. A variety of non-standard event types recorded and typo make up wide range of event types. The values stored in the evtype column were corrected to fit the standard 48 through cross referencing the NOAA storm database analysis handbook.

```
stormdt$evtype <- tolower(stormdt$evtype)
stormdt$evtype <- trimws(stormdt$evtype, which = "both")
stormdt$evtype <- gsub("  ", " ", stormdt$evtype)
stormdt$evtype <- gsub("^tstm wind.*", "thunderstorm wind", stormdt$evtype)
stormdt$evtype <- gsub("^thunderstorm.*", "thunderstorm wind",
                       stormdt$evtype)
stormdt$evtype <- gsub("^hurricane.*", "hurricane(typhoon)",
                       stormdt$evtype)
stormdt$evtype <- gsub("^typhoon.*", "hurricane(typhoon)", stormdt$evtype)
stormdt$evtype <- gsub("^high wind.*", "high wind", stormdt$evtype)
stormdt$evtype <- gsub("^gust.*", "marine high wind", stormdt$evtype)
stormdt$evtype <- gsub("^non.*", "high wind", stormdt$evtype)
stormdt$evtype <- gsub(".*fire.*", "wildfire", stormdt$evtype)
stormdt$evtype <- gsub(".*surf.*", "high surf", stormdt$evtype)
stormdt$evtype <- gsub(".*astronomical high tide.*", "high surf",
                       stormdt$evtype)
stormdt$evtype <- gsub(".*coastal flooding.*", "coastal flood",
                       stormdt$evtype)
stormdt$evtype <- gsub("gradient wind", "tropcial depression",
                       stormdt$evtype)
```

```
stormdt$evtype <- gsub("landspout", "dust devil", stormdt$evtype)
stormdt$evtype <- gsub("lake effect snow", "lake-effect snow",
                       stormdt$evtype)
stormdt$evtype <- gsub("tropcial depression", "tropical depression",
                       stormdt$evtype)
stormdt$evtype <- gsub("marine tstm wind", "marine thunderstorm wind",
                       stormdt$evtype)
stormdt$evtype <- gsub("glaze", "freezing fog", stormdt$evtype)
stormdt$evtype <- gsub("^tropical storm wind.*", "tropical storm",
                       stormdt$evtype)
stormdt$evtype <- gsub(".*rain.*", "heavy rain", stormdt$evtype)
stormdt$evtype <- gsub(".*microburst.*", "thunderstorm wind",
                       stormdt$evtype)
stormdt$evtype <- gsub(".*whirlwind.*", "thunderstorm wind",
                       stormdt$evtype)
stormdt$evtype <- gsub(".*downburst.*", "thunderstorm wind",
                       stormdt$evtype)
stormdt$evtype <- gsub(".*small hail.*", "hail", stormdt$evtype)
stormdt$evtype <- gsub(".*blowing dust.*", "dust storm", stormdt$evtype)
stormdt$evtype <- gsub(".*fog.*", "dense fog", stormdt$evtype)
stormdt$evtype <- gsub(".*coastalstorm.*", "tropical storm",
                       stormdt$evtype)
stormdt$evtype <- gsub(".*coastal storm.*", "tropical storm",
                       stormdt$evtype)
stormdt$evtype <- gsub(".*strong winds.*", "strong wind", stormdt$evtype)
stormdt$evtype <- gsub(".*heavy snow shower.*", "heavy snow",
                       stormdt$evtype)
stormdt$evtype <- gsub(".*storm surge.*", "storm surge/tide",
                       stormdt$evtype)
stormdt$evtype <- gsub(".*warm weather.*", "heat", stormdt$evtype)
stormdt$evtype <- gsub(".*winds.*", "high wind", stormdt$evtype)
stormdt$evtype <- gsub("^wind$", "high wind", stormdt$evtype)
stormdt$evtype <- gsub(".*excessive snow*", "blizzard", stormdt$evtype)
stormdt$evtype <- gsub("^snow$", "heavy snow", stormdt$evtype)
stormdt$evtype <- gsub(".*late season snow.*", "heavy snow",
                       stormdt$evtype)

flood <- c("flood", "dam", "ice jam", "fld", "river flood",
           "lakeshore flood", "river flooding", "high water")
for(f in flood){stormdt$evtype <- gsub(paste(".*", f, ".*", sep=""),
                                       "flood", stormdt$evtype)}

frost <- c("freeze", "hard freeze", "agricultural freeze", "frost")
for(f in frost){stormdt$evtype <- gsub(paste(".*", f, ".*", sep=""),
                                     "frost/freeze", stormdt$evtype)}

winter <- c("freezing", "black ice", "icy", "ice roads",
                   "ice on road", "light snow", "snow squall", "wintry",
                   "winter", "snow and ice", "rain/snow", "cold and snow",
                   "mixed precip", "falling snow/ice", "blowing snow")

for(w in winter){stormdt$evtype <- gsub(paste(".*", w, ".*", sep=""),
                                              "winter weather", stormdt$evtype)}
```

```
heat <- c("heat wave", "record heat", "record excessive heat",
          "hyperthermia", "unseasonably warm")
for(h in heat){stormdt$evtype <- gsub(paste(".*", h, ".*", sep = ""),
                                      "excessive heat", stormdt$evtype)}

xcold <- c("extreme cold", "unseasonable cold", "extended cold",
           "unseasonably cold", "hypothermia", "extreme windchill")
for(c in xcold){stormdt$evtype <- gsub(paste(".*", c, ".*", sep = ""),
                                      "extreme cold/wind chill",
                                      stormdt$evtype)}

cold <- c("cold", "cold temperature", "cold weather")
for(c in cold){stormdt$evtype <- gsub(paste(".*", c, ".*", sep=""),
                                      "cold/wind chill", stormdt$evtype)}

seas <- c("heavy seas", "high seas", "rough seas", "rip currents",
          "wind and wave", "rogue wave", "high swells", "marine accident")
for(s in seas){stormdt$evtype <- gsub(paste(".*", s, ".*", sep=""),
                                      "high surf",
                                      stormdt$evtype)}

debris <- c("mudslide", "mud slide", "mudslides", "rock slide",
            "landslide","landslides", "erosion","landslump", "debris")
for(d in debris){stormdt$evtype <- gsub(paste(".*", d, ".*", sep = ""),
                                        "debris flow", stormdt$evtype)}
```

Decision was made to eliminate evtype values that did not have clear relationship to standard 48; these events("other" amd "drowning") were only present in 1 record respectively.

```
stormdt <- stormdt[stormdt$evtype != "other" & stormdt$evtype != "drowning",]
```

For presentation's sake, names of event types are capitalized.

```
evtype_capitalize <- function(x) {
        s <- strsplit(x, " ")[[1]]
        paste(toupper(substring(s, 1,1)), substring(s, 2), sep="", collapse=" ")
}

stormdt$evtype <- sapply(stormdt$evtype, evtype_capitalize)
```

Analysis

All visualizations for storm data will be generated using the ggplot2 package from Hadley Wickman, begining with a histogram showing the event type counts for the time frame of January 1, 1996 to November 30, 2011.

```
library(ggplot2);
evtype_hist <- qplot(stormdt$evtype,
                     main = "Event Type Frequency accross USA",
                     xlab = "Event Type",
                     ylab = "Count")
evtype_hist +
        theme(axis.text.x = element_text(angle = 45, hjust = 1),
              plot.title = element_text(hjust = 0.5))
```

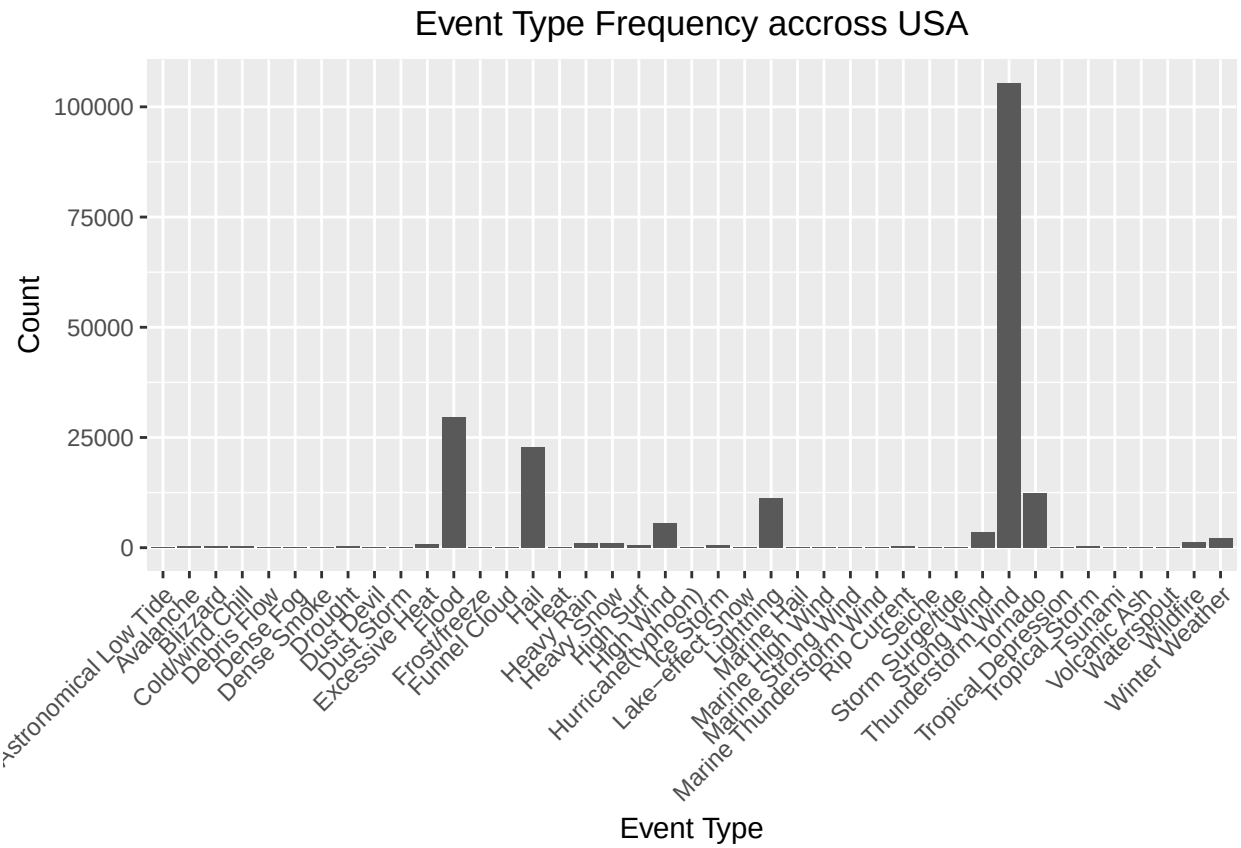

Table reflecting the values of top 6 events by count:

```
knitr::kable(head(count(stormdt, evtype) %>% arrange(desc(n))),
             caption = "Event Counts 1996-2011", col.names = c("Event", "Count"))
```

Table 1: Event Counts 1996-2011

Event	Count
Thunderstorm Wind	105453
Flood	29527
Hail	22690
Tornado	12365
Lightning	11151
High Wind	5476

####Examining impact on health by event type

For the purposes of calculation, a timeframe variable is created using the difftime function with units set to "weeks", the value of function then divded by 52.25 (accounting for the impact of leap years) to set the unit it years. The returned value is then coerced to a numeric value for computation.

A new data.table is created by grouping the data by event type, creating summary statistics for total fatalities, mortatity by event occurence, number of occurences, and finally occurences of a given event type by year (calculated with timeframe variable). ggplot() is employed to visualize total fatalities by event.

```
timeframe <- as.numeric(difftime(max(stormdt$bgn_date), min(stormdt$bgn_date),
                        units = "weeks") / 52.25)
```

```
event_fatalities <- stormdt %>%
        group_by(evtype) %>%
        summarise(totalfatalities = sum(fatalities),
                  fatalitiesper = (totalfatalities / n()),
                  occurences = n(),
                  occurencesperyear = occurences / timeframe) %>%
        arrange(desc(totalfatalities))

ggplot(event_fatalities, aes(evtype, totalfatalities)) +
        geom_bar(stat = "sum") +
        labs(title = "Fatalities by Event Type",
             x = "Event Type", y = "Total Fatalities") +
        theme(axis.text.x = element_text(angle = 45, hjust = 1),
              plot.title = element_text(hjust = 0.5),
              legend.position = "none")
```

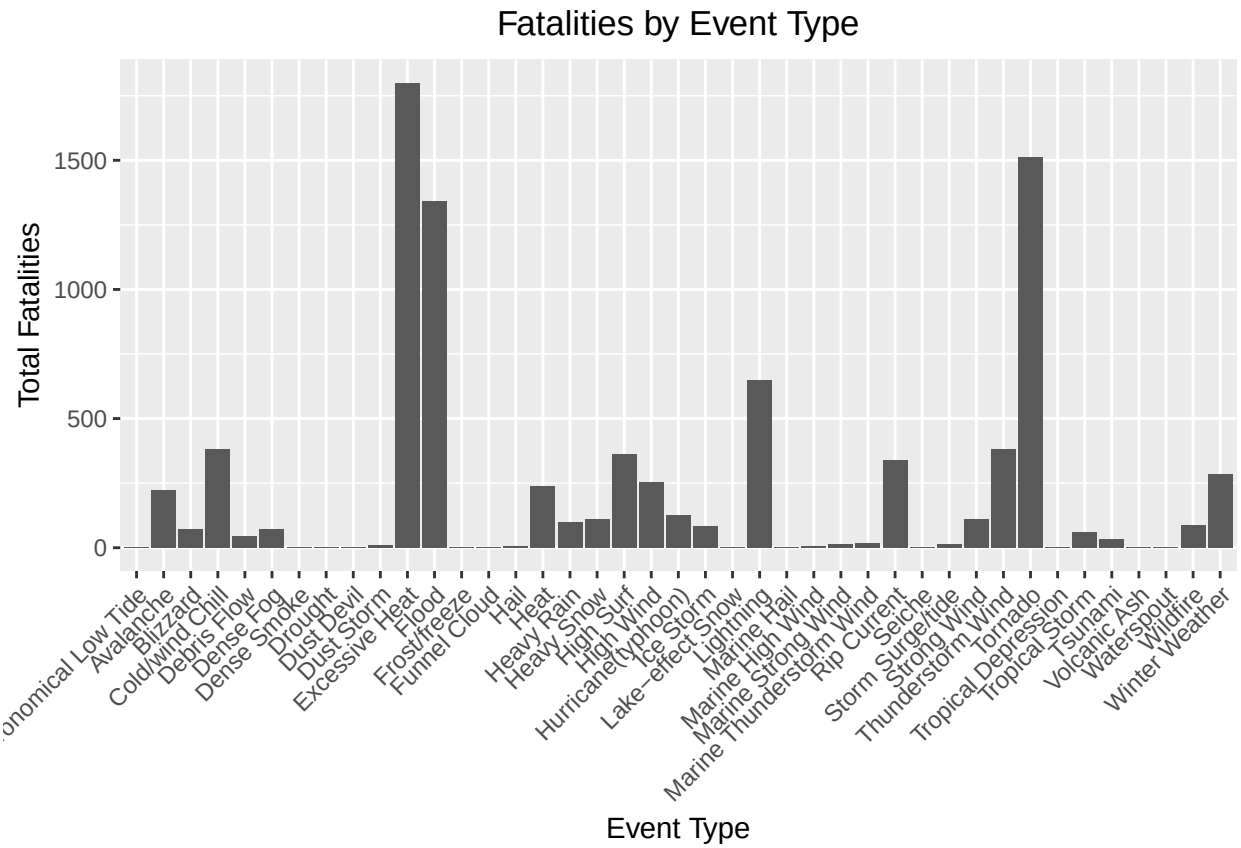

Table below expands on the data visualized above:

```
knitr::kable(head(event_fatalities), caption = "Total Fatalities by Event",
             col.names = c("Event", "Total Fatalities",
                           "Mortality (by occurence)", "Occurences",
                           "Yearly rate of occurence"))
```

Table 2: Total Fatalities by Event

Event	Total Fatalities	Mortality (by occurence)	Occurences	Yearly rate of occurence
Excessive Heat	1800	2.5974026	693	NA

Event	Total Fatalities	Mortality (by occurence)	Occurences	Yearly rate of occurence
Tornado	1511	0.1221998	12365	NA
Flood	1340	0.0453822	29527	NA
Lightning	650	0.0582907	11151	NA
Thunderstorm Wind	383	0.0036319	105453	NA
Cold/wind Chill	382	0.9095238	420	NA

By the visualization and table above, t he 3 event types which have the greatest impact of mortality are: Excessive

Excessive Heat - most impacted states - count by month

```
excessive_heat <- stormdt[which(stormdt$evtype == "Excessive Heat"),]
nrow(excessive_heat)
```

```
## [1] 693
```

Tornado

-summary stats: duration, widtch, avg damage by F type, avg damage by width, avg damage by length, state with most torandos(damage by year) -visualizaitons:

```
tornado <- stormdt[which(stormdt$evtype == "Tornado"),]
nrow(tornado)
```

```
## [1] 12365
```

Flood

```
flood <- stormdt[which(stormdt$evtype == "Flood"),]
nrow(flood)
```

```
## [1] 29527
```

Examining Injuries by event type:

Summary dataset is generated through grouping data by event type, summarising total injuries, injuries sustained by occurence of event, number of event occurences, and occurences by year. ggplot2 is used to visualize the total number of injuries by event type.

```
event_injuries <- stormdt %>%
        group_by(evtype) %>%
        summarise(totalinjuries = sum(injuries),
                  injuryperoccurence = totalinjuries / n(),
                  occurences = n(),
                  occurencesperyear = occurences/timeframe) %>%
        arrange(desc(totalinjuries))

ggplot(event_injuries, aes(evtype, totalinjuries)) +
        geom_bar(stat = "sum") +
        labs(title = "Injuries by Event Type since 1996",
             x = "Event Type", y = "Total Injuries") +
        theme(axis.text.x = element_text(angle = 45, hjust = 1),
              plot.title = element_text(hjust = 0.5),
              legend.position = "none")
```

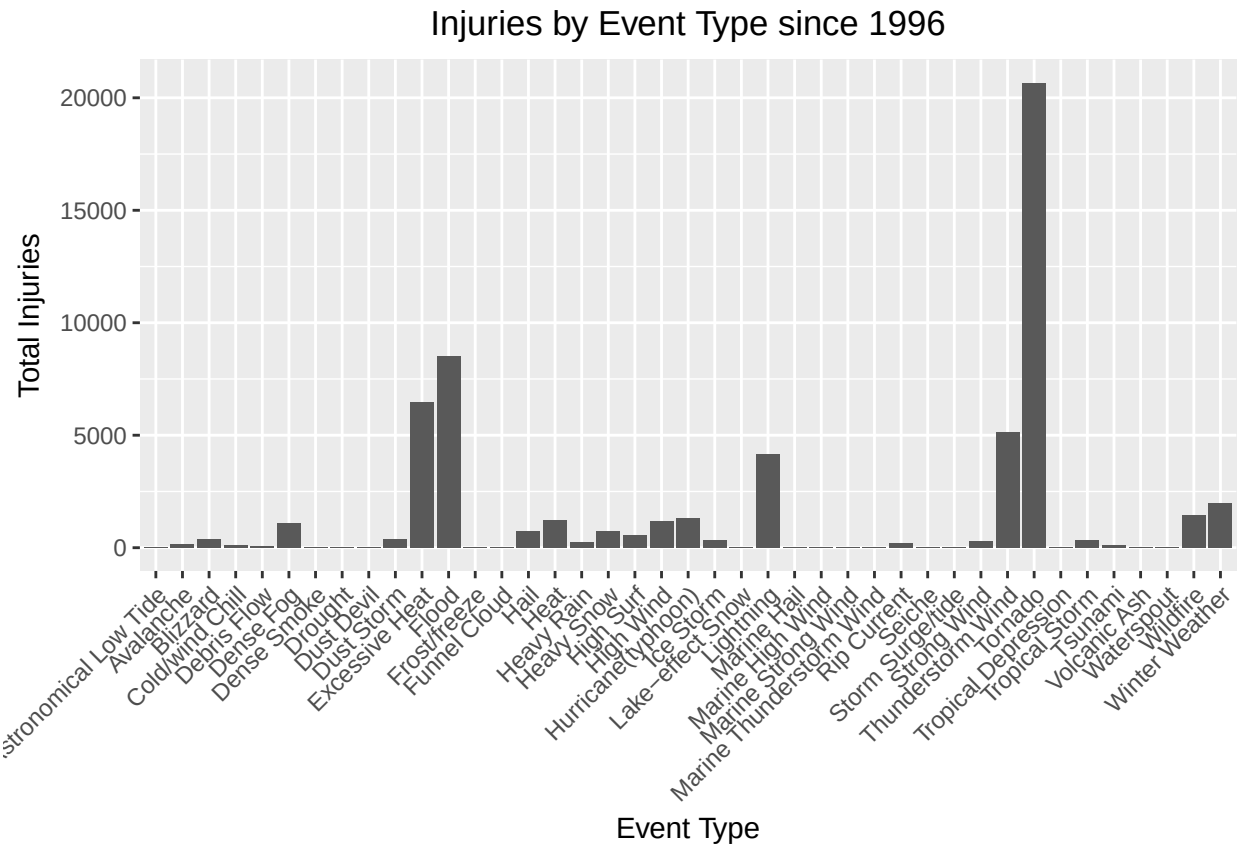

Table below expands on the data visualized above:

```
knitr::kable(head(event_injuries), caption = "Total Injuries by Event since 1996",
             col.names = c("Event", "Total Injuries", "Injuries per Occurence",
                           "Occurences", "Occurences per Year"))
```

Table 3: Total Injuries by Event since 1996

Event	Total Injuries	Injuries per Occurence	Occurences	Occurences per Year
Tornado	20667	1.6714112	12365	NA
Flood	8520	0.2885495	29527	NA
Excessive Heat	6478	9.3477633	693	NA
Thunderstorm Wind	5154	0.0488749	105453	NA
Lightning	4140	0.3712672	11151	NA
Winter Weather	1977	0.8925508	2215	NA
#### Economic Impact				

Property Damage

```
event_propdmg <- stormdt %>%
        group_by(evtype) %>%
        summarise(totalpropdmg = sum(propdmg),
                  damageperoccurence = totalpropdmg / n(),
                  occurences = n(),
                  occurencesperyear = occurences/timeframe) %>%
        mutate(totalpropdmg = totalpropdmg / 1E9,
```

```
                damageperoccurence = damageperoccurence / 1E6) %>%
        arrange(desc(totalpropdmg))

ggplot(event_propdmg, aes(evtype, totalpropdmg)) +
        geom_bar(stat = "sum") +
        labs(title = "Total Property Damage by Event Type since 1996",
             x = "Event Type", y = "Total Damage (in Billions)") +
        theme(axis.text.x = element_text(angle = 45, hjust = 1),
              plot.title = element_text(hjust = 0.5),
              legend.position = "none")
```

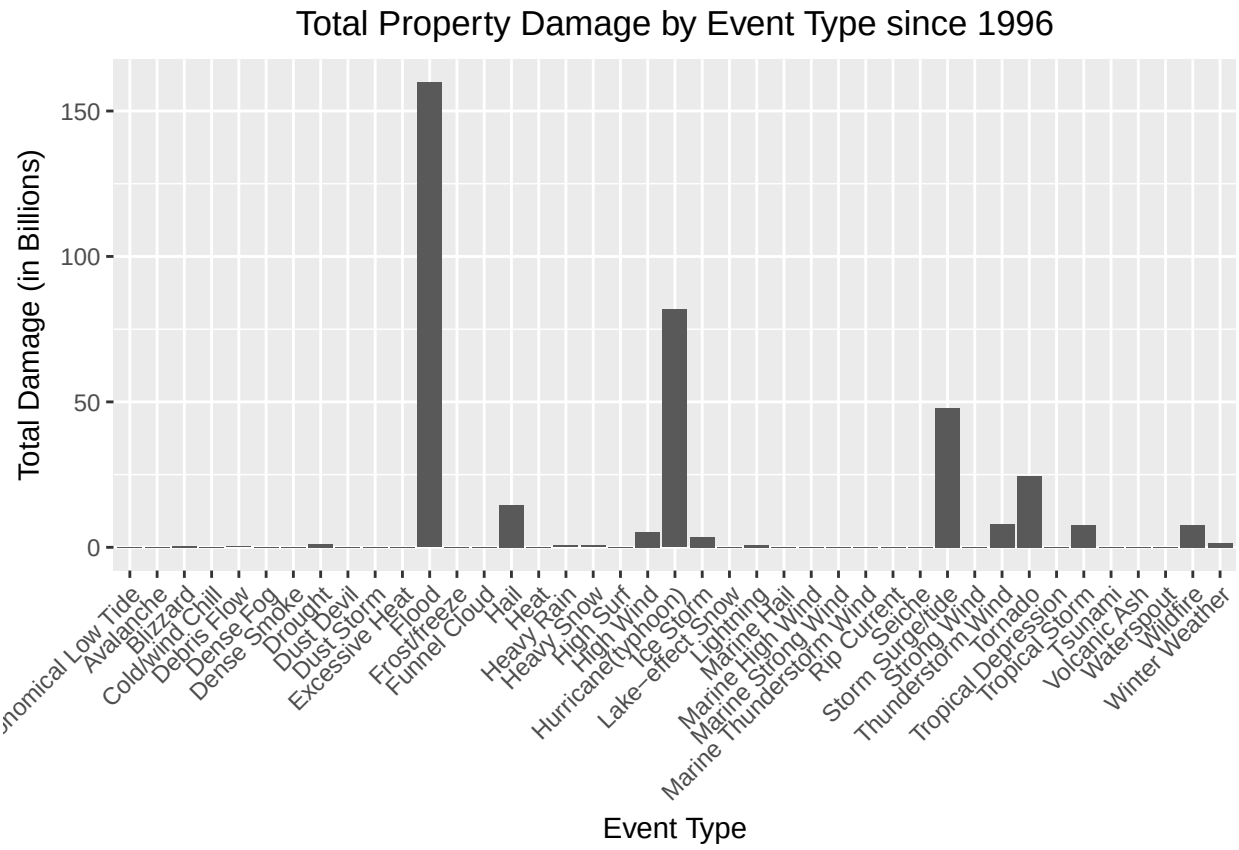

Table below expands upon property damage by event type

```
knitr::kable(head(event_propdmg), caption = "Property Damage by Event since 1996",
             col.names = c("Event", "Damage (in Billions)",
                           "Average Damage (in Millions)","Occurences",
                           "Occurences per Year"))
```

Table 4: Property Damage by Event since 1996

Event	Damage (in Billions)	Average Damage (in Millions)	Occurences	Occurences per Year
Flood	159.774855	5.4111442	29527	NA
Hurricane(typhoon)	81.718889	392.8792741	208	NA
Storm Surge/tide	47.834724	221.4570556	216	NA
Tornado	24.616906	1.9908537	12365	NA
Hail	14.595213	0.6432443	22690	NA

Event	Damage (in Billions)	Average Damage (in Millions)	Occurences	Occurences per Year
Thunderstorm Wind	7.915343	0.0750604	105453	NA

Flood

#####Hurricane

#####Storm Surge

Crop Damage

```
event_cropdmg <- stormdt %>%
        group_by(evtype) %>%
        summarise(totalcropdmg = sum(cropdmg),
                  cropdmgperoccurence = totalcropdmg/n(),
                  occurences = n(),
                  occurencesperyear = occurences / timeframe) %>%
        mutate(totalcropdmg = totalcropdmg / 1E9,
               cropdmgperoccurence = cropdmgperoccurence/1E6) %>%
        arrange(desc(totalcropdmg))

ggplot(event_cropdmg, aes(evtype, totalcropdmg), fill = evtype) +
        geom_bar(stat = "sum") +
        labs(title = "Total Crop Damage by Event Type",
             x = "Event Type", y = "Total Damage (Billions)") +
        theme(axis.text.x = element_text(angle = 45, hjust = 1),
              plot.title = element_text(hjust = 0.5),
              legend.position = "none")
```

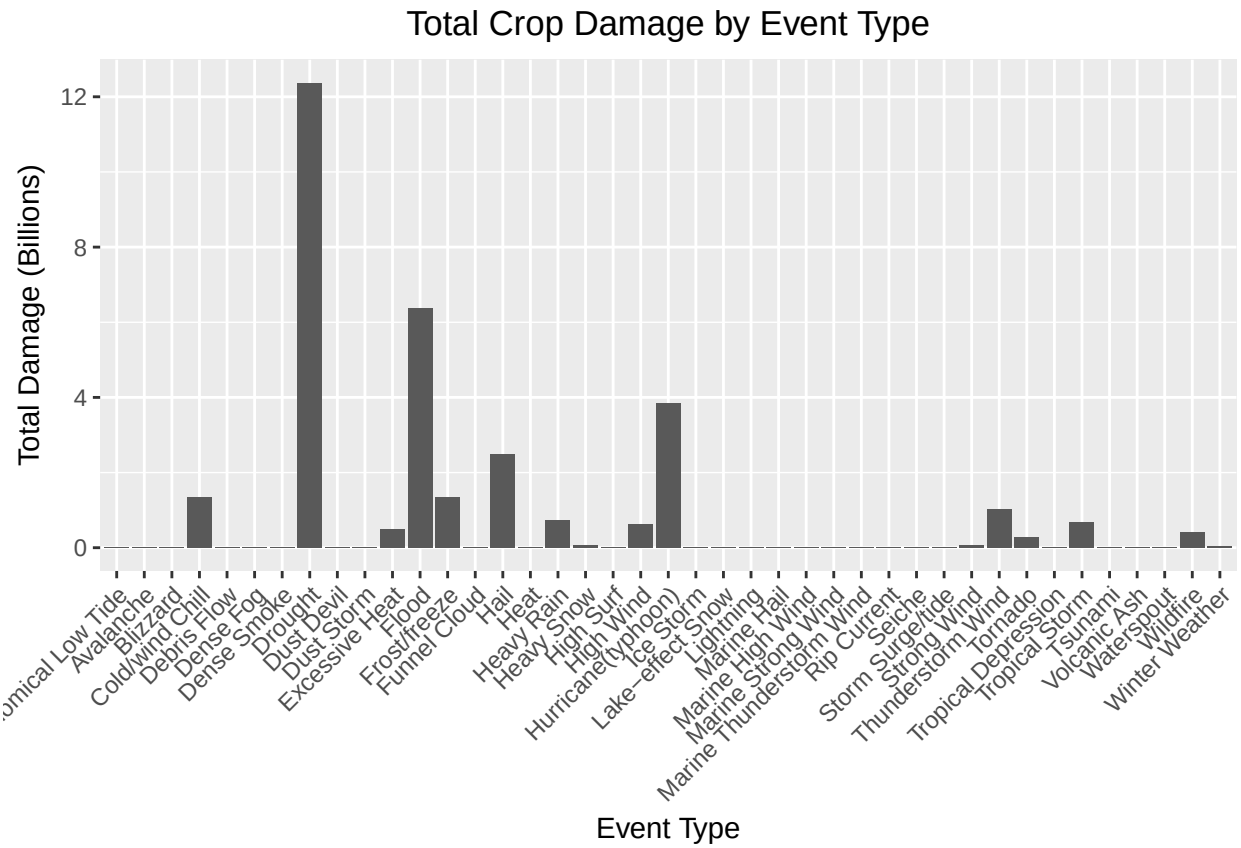

Table below expands upon crop damage.

```
knitr::kable(head(event_cropdmg), caption = "Crop Damage by Event",
             col.names = c("Event", "Total Damage (in Billions)",
                           "Damage per Occurence (in Millions)", "Occurences",
                           "Occurences per Year"))
```

Table 5: Crop Damage by Event

Event	Total Damage (in Billions)	Damage per Occurence (in Millions)	Occurences	Occurences per Yea
Drought	12.367566	47.9363023	258	NA
Flood	6.382193	0.2161477	29527	NA
Hurricane(typhoon)	3.840108	18.4620567	208	NA
Hail	2.496822	0.1100407	22690	NA
Cold/wind Chill	1.356766	3.2303940	420	NA
Frost/freeze	1.334631	9.7418321	137	NA

Drought

Flood

Hurricane

Results