NOAA weather health and property risk analysis

Synopsis

This paper is intended to demonstrate the most common causes of injury/death and property damage due to weather events. This was determined by calculating the number of events that resulted in injury or death and ordering these events based on the number of occurences of said event. Property damage was determined based on the number of events that totalled over 1 Billion USD in damages.

Data Processing

Downloading

First: we downloaded the data from the website listed below. Only the sections of interest were loaded to reduce the amount of processing power required. Code is shown below:

```
library(tidyverse)
#download.file("https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2", "StormData.csv.bz2")
raw_Data <- read.csv("StormData.csv.bz2")
raw_Data <- raw_Data[,c("EVTYPE","FATALITIES","INJURIES","PROPDMG","PROPDMGEXP")]
```

Renaming

Many similar events were listed in the report but with differing event types. In order to calculate the number of occurences of the events properly, it was determined that a single event name should be used for all similar event types. The renaming of the events are shown below:

```
raw_Data$EVTYPE <- str_replace_all(raw_Data$EVTYPE, "HEAT WAVE", "EXCESSIVE HEAT")
raw_Data$EVTYPE <- str_replace_all(raw_Data$EVTYPE, "RIP CURRENTS", "RIP CURRENT")
raw_Data$EVTYPE <- str_replace_all(raw_Data$EVTYPE, "HEAVY SNOW", "WINTER STORM")
raw_Data$EVTYPE <- str_replace_all(raw_Data$EVTYPE, "STRONG WIND", "HIGH WIND")
raw_Data$EVTYPE <- str_replace_all(raw_Data$EVTYPE, "EXTREME COLD/WIND CHILL", "WINTER STORM")
raw_Data$EVTYPE <- str_replace_all(raw_Data$EVTYPE, "BLIZZARD", "WINTER STORM")
raw_Data$EVTYPE <- str_replace_all(raw_Data$EVTYPE, "EXTREME HEAT", "EXCESSIVE HEAT")
raw_Data$EVTYPE <- str_replace_all(raw_Data$EVTYPE, "COLD/WIND CHILL", "WINTER STORM")
raw_Data$EVTYPE <- str_replace_all(raw_Data$EVTYPE, "ICE STORM", "WINTER STORM")
raw_Data$EVTYPE <- str_replace_all(raw_Data$EVTYPE, "TSTM WINDS", "TSTM WIND")

raw_Data$EVTYPE <- gsub(".*HURRICANE.*", "HURRICANE", raw_Data$EVTYPE, perl = TRUE)
raw_Data$EVTYPE <- gsub( ".*HIGH SURF.*", "HIGH SURF", raw_Data$EVTYPE, perl = TRUE)
raw_Data$EVTYPE <- gsub( ".*THUNDERSTORM.*", "TSTM WIND", raw_Data$EVTYPE, perl = TRUE)
raw_Data$EVTYPE <- gsub( ".*HEAT.*", "EXCESSIVE HEAT", raw_Data$EVTYPE, perl = TRUE)
raw_Data$EVTYPE <- gsub( ".*STORM SURGE.*", "HURRICANE", raw_Data$EVTYPE, perl = TRUE)
```

Calculating most common events for fatalities

The next step was to calculate the most common events that resulted in at least one fatality. These events were then split off into a list which would be used to subset the raw data for creating the box plot. Code is shown below:

```
sum_health_Data <- aggregate(cbind(raw_Data$FATALITIES, raw_Data$INJURIES),
    by=list(raw_Data$EVTYPE), FUN=sum) %>%
    subset(V1 != 0 & V2 != 0)

names(sum_health_Data) <- c("Event", "Fatalities", "Injuries")
sum_health_Data <- sum_health_Data[order(-sum_health_Data$Fatalities),]
event_type_fatalities <- sum_health_Data$Event[1:6]
```

Calculating most common events for injuries

This step is nearly identical to the one above. However, injury data has been swapped in place of the fatality data. This allows for events with large numbers of injuries but few fatalities to be found. Code is shown below:

```
sum_health_Data <- aggregate(cbind(raw_Data$FATALITIES, raw_Data$INJURIES),
    by=list(raw_Data$EVTYPE), FUN=sum) %>%
    subset(V1 != 0 & V2 != 0)
names(sum_health_Data) <- c("Event", "Fatalities", "Injuries")
sum_health_Data <- sum_health_Data[order(-sum_health_Data$Injuries),]
event_type_inujuries <- sum_health_Data$Event[1:6]
```

Calculating most common events with Billion dollar property damages

It was decided to use only events with damages totalling in the billions. This was because if a single event totalled 1 Million in property damage, it would take 1,000 of these events to equal one event with damages totalling over 1 Billion. The code for this calculation is shown below:

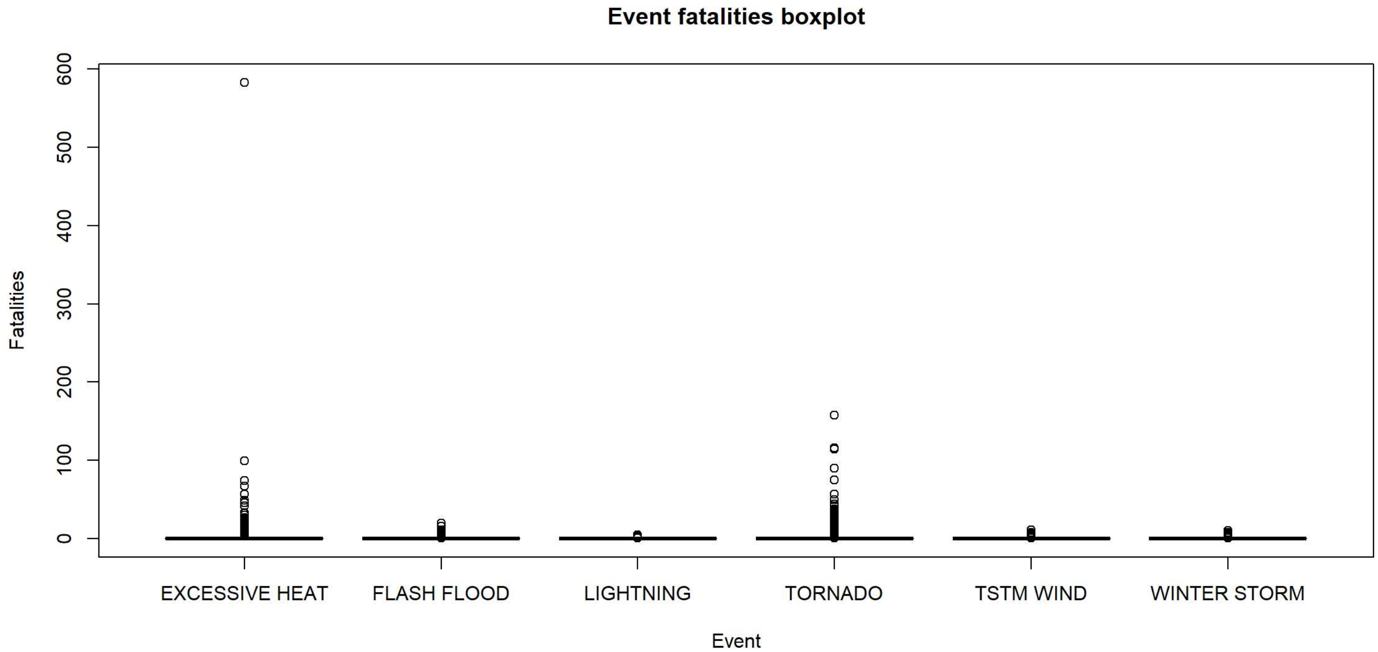
```
cost_data <- table(subset(raw_Data, PROPDMGEXP == "B"))
```

Results

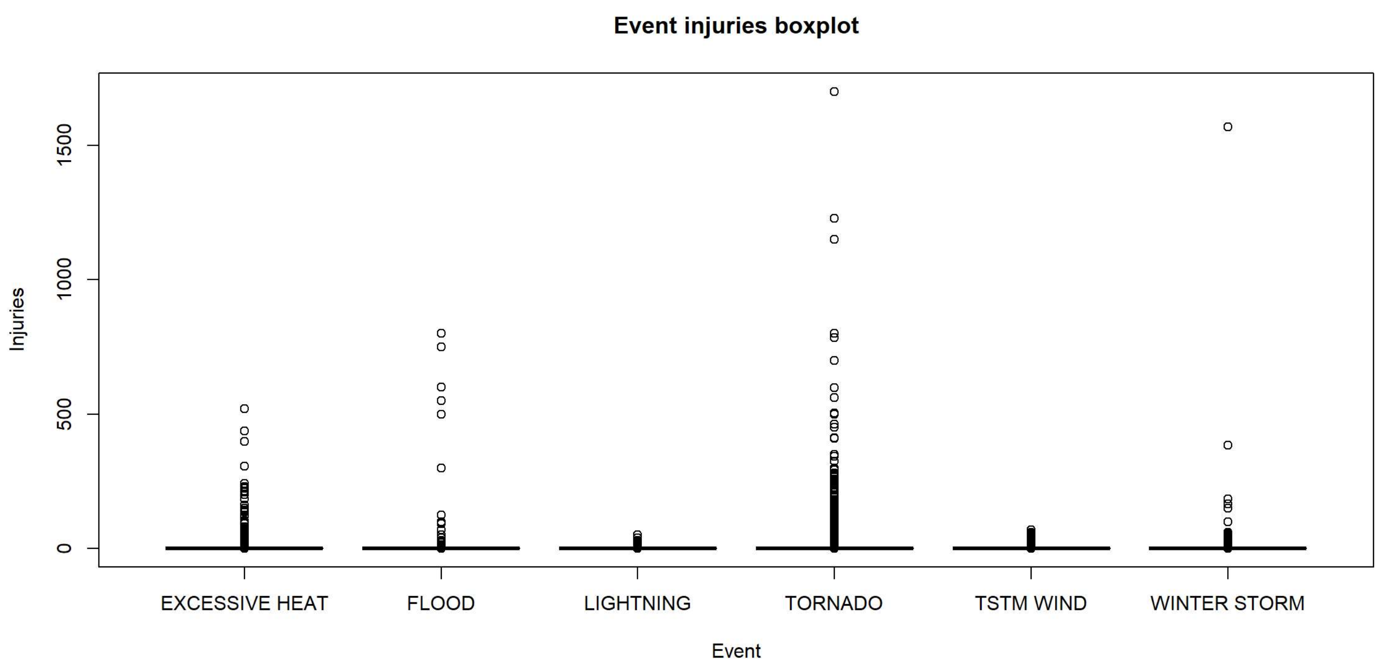
Health related

From the data listed above, the following graphs were produced for fatalities and injuries respectively:

```
top_6_data <- subset(raw_Data, EVTYPE %in% event_type_fatalities)
boxplot(top_6_data$FATALITIES ~ top_6_data$EVTYPE, xlab = "Event",
    ylab = "Fatalities", main = "Event fatalities boxplot")
```




```
top_6_data <- subset(raw_Data, EVTYPE %in% event_type_inujuries)
boxplot(top_6_data$INJURIES ~ top_6_data$EVTYPE, xlab = "Event",
    ylab = "Injuries", main = "Event injuries boxplot")
```

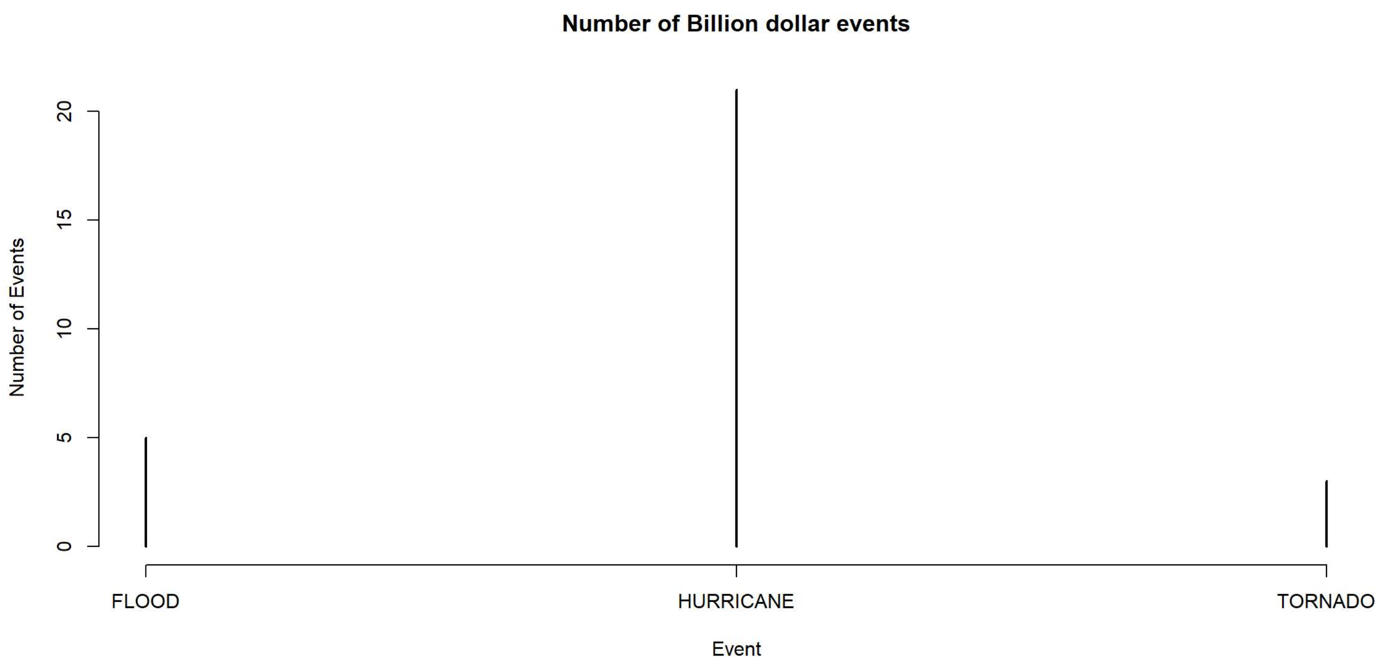


The data shows that of all the weather events listed above, Tornados have the greatest risk for causing large numbers of injuries and fatalities. Excessive heat also shows a high number of injuries and fatalities, including one outlier event that surpased all other events and thus should also be taken into consideration when considering disaster preparedness. Flooding and winter storms show a high likelyhood of causing large numbers of injuries but fatalities are relatively rarer for these events, likely due to the fact that these events often occur in the same areas and thus the local populations are ready for events of these types.

Property damage

From the processing listed previously the following graph was generated:

```
plot(cost_table[cost_table>1], type = "h", xlab = "Event",
    ylab = "Number of Events", main = "Number of Billion dollar events")
```



From the graph, it is clear that Hurricanes are the leading cause of Billion dollar weather-related disasters, with 21 Billion dollar disaster events being recorded. From this data, it is apparent that any area in the path of Hurricanes should perpare for large property damage events.