Consequences of Severe Weather on US Population

REK

06/04/2018

Synopsis

For the cost to public health, **Tornados** are the most dangerous in terms of absolute number of fatalities and injuries. However, such events are frequent and, on average, only 1 death occurs every 10 tornado events despite there being an average of nearly 2 injuries to eveny event. **Tsunamis** on the other hand, while having a low total number of deaths, are the most deadly event when they occur, killing at least 1 person and injuring an average of 6 people per event.

For the cost to property and crop damage; property damage was significantly greater than that to crops. Such property damage was largely caused by **Floods** and **Hurricanes** with 150.483 Billion USD and 85.356 billion USD respectivly. In terms of crop damage however, this was unuprisingly largely due to **Droughts** and **Floods**, causing 13.973 bilion USD and 10.956 billion USD worth of damage respectivly.

Data Processing and Cleaning

1. Load data and rename entries with multiple weather events
2. Sort weather events using clear naming conventions
3. Calculate the actual

1) Firstly we want to load the data set and clean the data up. For the event type variable, not all were inputted using the correct naming criteria so we want to clean this data up to allow a repliable analysis. Specifically here, some entries contained multiple different event types - in these cases, we will only take the first entry inputted.

```r
# Load in relevant R packages
library(dplyr)
library(ggplot2)
library(reshape2)
# Set the number format of output numbers
options(scipen = 99999, digits = 3)

# Set the working directory to the data file location
wd <- "/Users/rek514/Documents/Data_science"
setwd(wd)
# Details on the location of the downloaded file
fileURL <- "https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2"
destfile <- "StormData.csv.bz2"

# If we havent downloaded the data yet, dowmlaod it now.
if (!file.exists(destfile)) {
    download.file(fileURL, destfile)
}

# Read in data set
dataset <- read.csv("StormData.csv.bz2", na.strings = "")
# Create a copy of the dataset to modify
```

```
clean_dataset <- dataset
# Convert the EVTYPE variable to a chavter rather then factor
clean_dataset$EVTYPE <- as.character(clean_dataset$EVTYPE)

# Some inputs have multiple event types so we want to remove
# them where possible This looks for seperaters (/, and, &)
# and only keep the first event type of this input.
problementry <- grep("/", clean_dataset$EVTYPE)
for (i in 1:length(problementry)) {
    clean_dataset$EVTYPE[(problementry[i])] <- gsub("/.*", "",
        clean_dataset$EVTYPE[(problementry[i])])
}
problementry2 <- grep("&", clean_dataset$EVTYPE)
for (i in 1:length(problementry2)) {
    clean_dataset$EVTYPE[(problementry2[i])] <- sub("&.*", "",
        clean_dataset$EVTYPE[(problementry[i])])
}
problementry3 <- grep("and", clean_dataset$EVTYPE, ignore.case = TRUE)
for (i in 1:length(problementry3)) {
    clean_dataset$EVTYPE[(problementry3[i])] <- sub("and.*",
        "", clean_dataset$EVTYPE[(problementry[i])])
}
```

2. As mentioned above, some inputs have not been inputted using the standard EV Type name, so we want to ensure all the same event type, regardless of exact naming convention, are analysed under the same category. For example 'Heavy Rain', 'Heavy Precipitation' and 'Torrential Rain' will all be categorised as 'Heavy Rain'.

This means that there will be some cases not used as they do not fit in which one of these predefined 48 weather events or it is unclear which category the event belongs to (e.g. 'Record high' may refer to heat, rain or wind). If the entry met either of these criteria, the case was removed from the analysis. This code will be included in the appendix found at the end of this page.

RESULTS

Across the United States, which types of weather event is most harmful to population health?

1. Which event causes the most overall fatalities and which has the greatest morality rate (averaged across each individual event)?

```
# Total fatality/injurt for each event type
type_grouped <- group_by(clean_dataset, Global_event)
type_data <- summarise(type_grouped,
                       Fatalities_total=sum(FATALITIES, na.rm = TRUE),
                       Fatalities_mean=mean(FATALITIES, na.rm = TRUE),
                       Injuries_total=sum(INJURIES, na.rm = TRUE),
                       Injuries_mean=mean(INJURIES, na.rm = TRUE))

# Order data by fatalities total with greatest first
type_data <- type_data[with(type_data, order(-Fatalities_total)), ]
```

```
#Melt the data
melted_fatalities <- melt((type_data[1:3]), id.vars = "Global_event")
# Remove rows with an event name of NA
melted_fatalities <- filter(melted_fatalities, !is.na(Global_event))

# Plots event types by total fatality
g <- ggplot(melted_fatalities, aes(Global_event,value, fill=factor(variable)))
g + geom_bar(stat='identity', position='dodge')+
  theme(axis.text.x = element_text(angle = 90, hjust = 1))+
  facet_grid(melted_fatalities$variable~., scales = 'free')+
  ggtitle('Total and Average Fatality rate from Weather events across the US')+
  xlab("Weather Event")+ylab("Total")+ theme(legend.position="none")
```

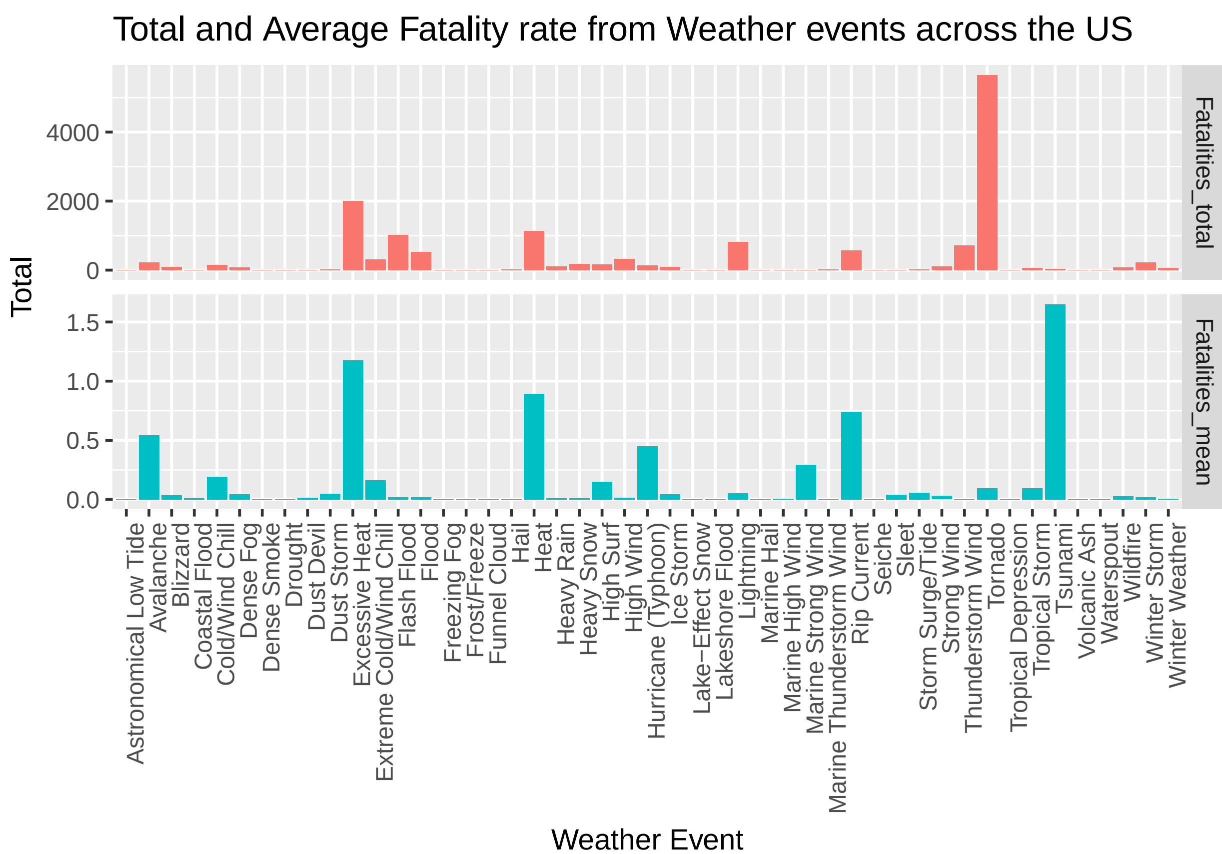


Figure 1: Figure 1: Bar graph to show the total number of fatalities from each weather event and the mean number of fatalities for each individual event.

The above graphs show that **Tornados** caused the highest number of fatalities in the US with 5658 deaths. This was followed by **Excessive Heat** and **Heat** which caused 1999 and 1131 deaths respectivly. However, when taking into consideration the average fatality rate (ratio of deaths to recorded weather events), **Tsunamis** are the most dangerous to public health with a average of 1.65 weather per event - in that sense, there is at least 1 death for every excessive heat event. **Excessive Heat** had the second highest fatality rate, again with atleast 1 death per event.

What about injury rates however?

```
# Order data by injuries total
type_data <- type_data[with(type_data, order(-Injuries_total)), ]

# Melt the data and remove NA rowns
melted_injuries <- melt((type_data[,-(2:3)]), id.vars = "Global_event")
melted_injuries <- filter(melted_injuries, !is.na(Global_event))

# Plot both the total and mean injury rates
g <- ggplot(melted_injuries, aes(Global_event,value, fill=factor(variable)))
g + geom_bar(stat='identity', position='dodge')+
  theme(axis.text.x = element_text(angle = 90, hjust = 1))+
  facet_grid(melted_fatalities$variable~., scales = 'free')+
  ggtitle('Total and Average Injury rate from Weather events across the US')+
  xlab("Weather Event")+ylab("Total")+ theme(legend.position="none")
```

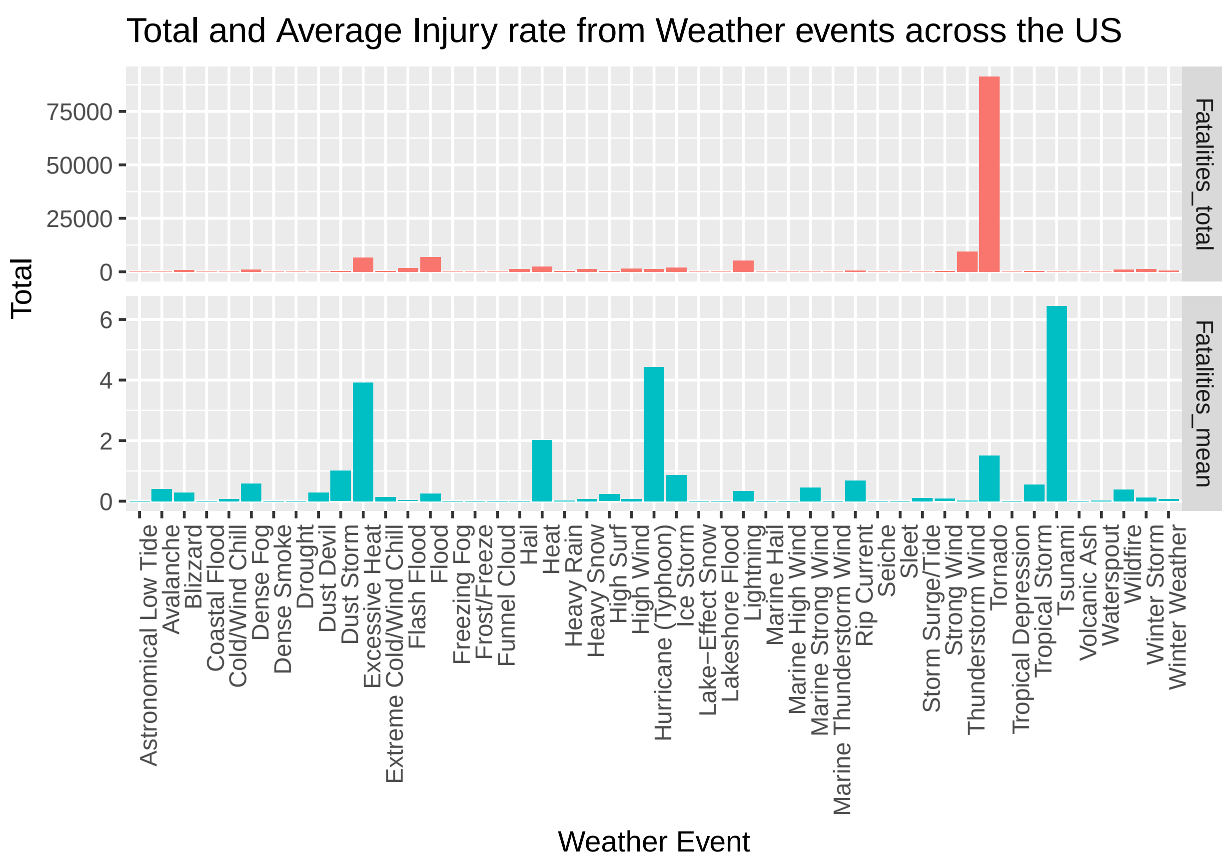


Figure 2: Figure 2: Bar graph to show the total number of injuries from each weather event and the mean number of injuries for each individual event

---

The above graphs also appear to show a similar story to the fatality data with **Tornados** caused the highest number of injuries in the US. However, both Thunder storms and Floods provide a greater danger to the public in terms of injuries compared to the heat events.

When looking at the average injury rate however, this shows a different picture. Here there is atleast one injury for every **Tsunami, Hurricane, Excessive Heat, Heat, Tornado** and **Dust storm** event. This is particually dangerous for **Tsunami** events where on average over 6 people are injured in every event.

RECOMMENDATION: Based on these findings, we would recommend that resourses should be put into the preduction and preparation for tornado events which are the most harmful to the US population whereby 1.5 people becomes injured on average in each event. However, such events tend to occur with high frequency and, when event occurance is taken into account, Tsunami and Heat events show both high fartality and injury rates.

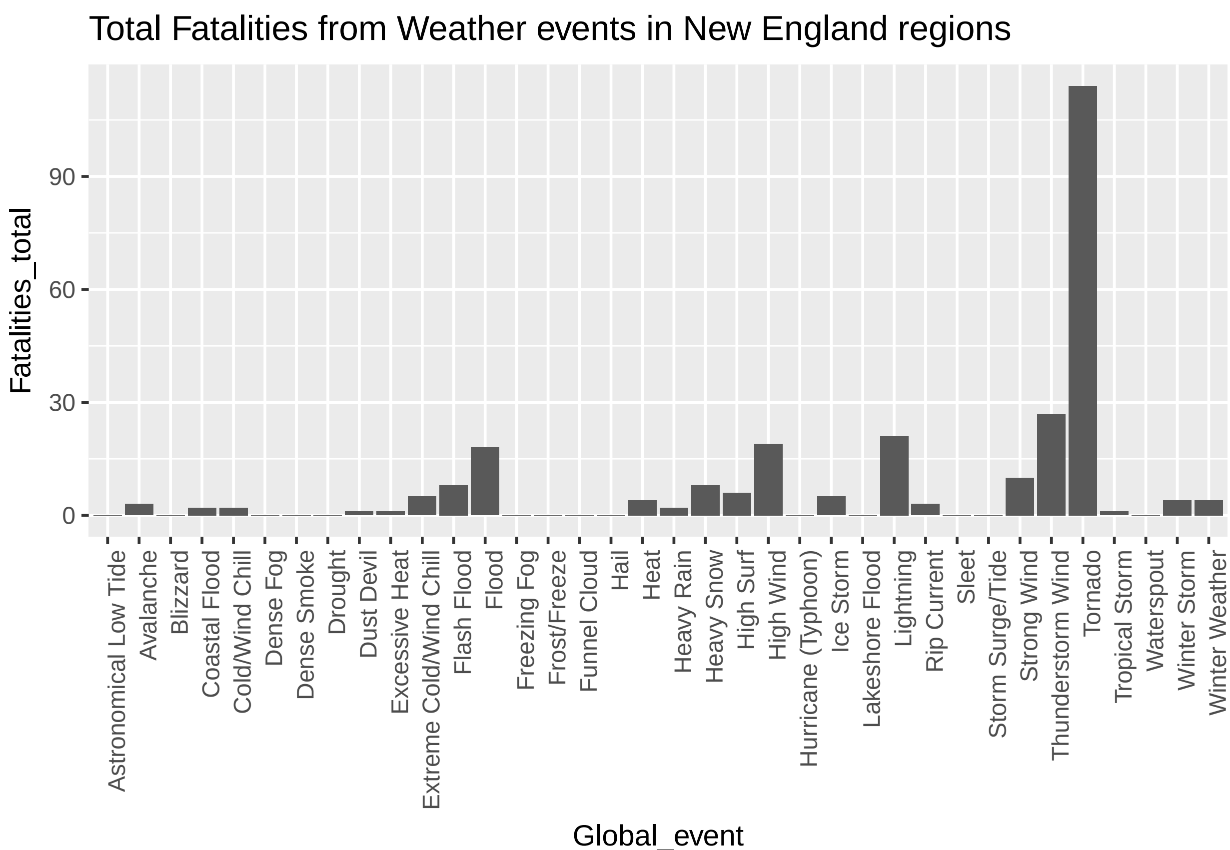
In that sense, as soon as authorities realise a Tsunami and Heat event is apparant, critical resourses should be put into preventing such problems.

What are the most dangerous weather events for the population of each US region?

Since the US has many weather climates, it may be beneficial to examine fatalities and injuries across smaller zones. The data here have been divided into each of the 9 US regions (Pacific, Mountain, West North Central, East North Central, West South Central, East South Central, New Endland, Middle Atlanitc and South Atlantic). Here we will just look at fatality rates.

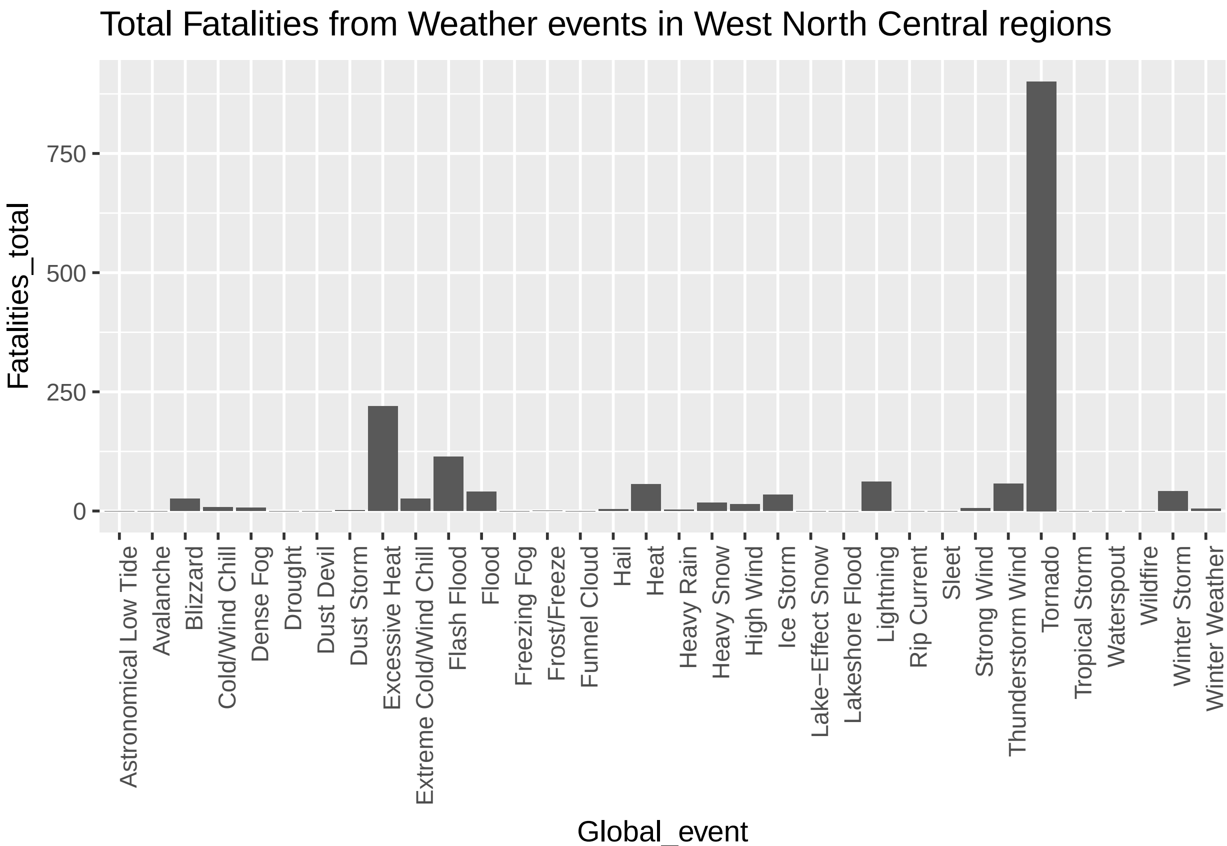
```
# Here we group the data by state and then by event type
state_type_grouped <- group_by(clean_dataset, STATE, Global_event)
state_type_data <- summarise(state_type_grouped,
                             Fatalities_total=sum(FATALITIES, na.rm = TRUE),
                             Fatalities_mean=mean(FATALITIES, na.rm = TRUE),
                             Injuries_total=sum(INJURIES, na.rm = TRUE),
                             Injuries_mean=mean(INJURIES, na.rm = TRUE),event_total=n())

# Orders the data in terms of total fatalities
state_type_data <- state_type_data[with(state_type_data, order(-Fatalities_total)), ]

# Defining state regions
state_regions <- vector(mode="list", length=9)
names(state_regions)<- c("New_England","Mid_Atlantic",
                         "East_North_Central","West_North_Central",
                         "South_Atlantic","East_South_Central",
                         "West_South_Central","Mountain","Pacific")

state_regions[['New_England']] <-c("CT", "MA", "ME", "NH", "RI", "VT")
state_regions[['Mid_Atlantic']] <-c("NY", "PA", "NJ")
state_regions[['East_North_Central']] <-c("WI", "MI", "OH","IN","IL")
state_regions[['West_North_Central']] <-c("ND", "SD", "NE","KS","MO","IA","MN")
state_regions[['South_Atlantic']] <-c("WV", "DC", "MD","DE","VA","NC","SC","GA","FL")
state_regions[['East_South_Central']] <-c("KY", "TN", "AL","MS")
state_regions[['West_South_Central']] <-c("TX", "OK", "AR","LA")
state_regions[['Mountain']] <-c("MT", "ID", "WY","NV","UT","CO","NM","AZ")
state_regions[['Pacific']] <-c("WA","OR","CA","HI","AK")

state_type_data[,8] <- NA
colnames(state_type_data)[8] <- 'Region'
```

```r
for (i in (1:(length(state_regions)))){
  this_region <- names(state_regions)[i]
  this_index <- grep(paste(state_regions[[i]], collapse = "|"), state_type_data$STATE)
  state_type_data$Region[this_index] <- this_region
  }

region_grouped <- group_by(state_type_data, Region, Global_event)
region_data <- summarise(region_grouped,
                         Fatalities_total=sum(Fatalities_total, na.rm = TRUE),
                         Fatalities_mean=mean(Fatalities_mean, na.rm = TRUE),
                         Injuries_total=sum(Injuries_total, na.rm = TRUE),
                         Injuries_mean=mean(Injuries_mean, na.rm = TRUE),
                         event_total=n())

region_data <- filter(region_data, !is.na(Global_event))
region_data <- filter(region_data, !is.na(Region))

# Tornado
new_england <- ggplot((region_data[grep("New_England", region_data$Region),]),
                      aes(Global_event, Fatalities_total))
new_england + geom_bar(stat='identity')+
  theme(axis.text.x = element_text(angle = 90, hjust = 1))+
  ggtitle('Total Fatalities from Weather events in New England regions')
```

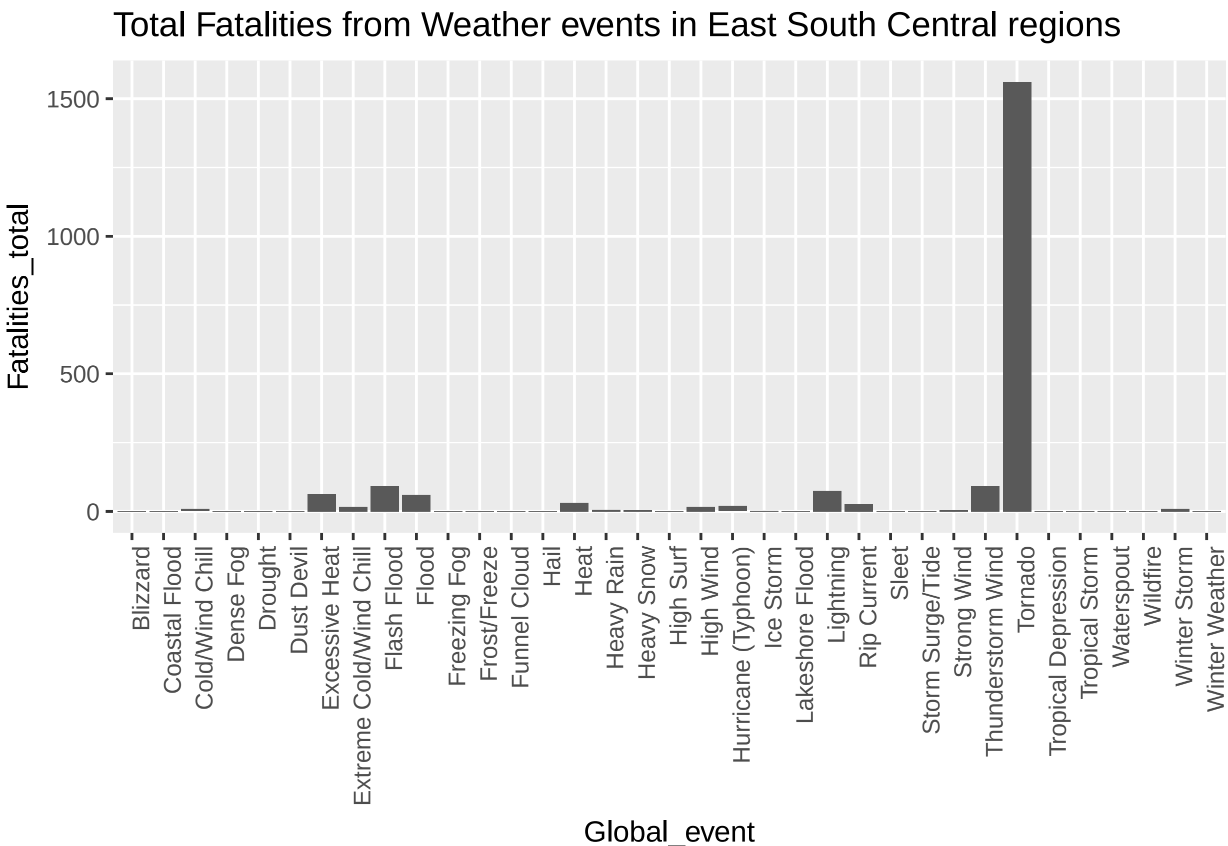



```r
# Tornado
wncentral<- ggplot((region_data[grep("West_North_Central", region_data$Region),]),
                   aes(Global_event, Fatalities_total))
```

```r
wncentral + geom_bar(stat='identity')+
  theme(axis.text.x = element_text(angle = 90, hjust = 1))+
  ggtitle('Total Fatalities from Weather events in West North Central regions')
```

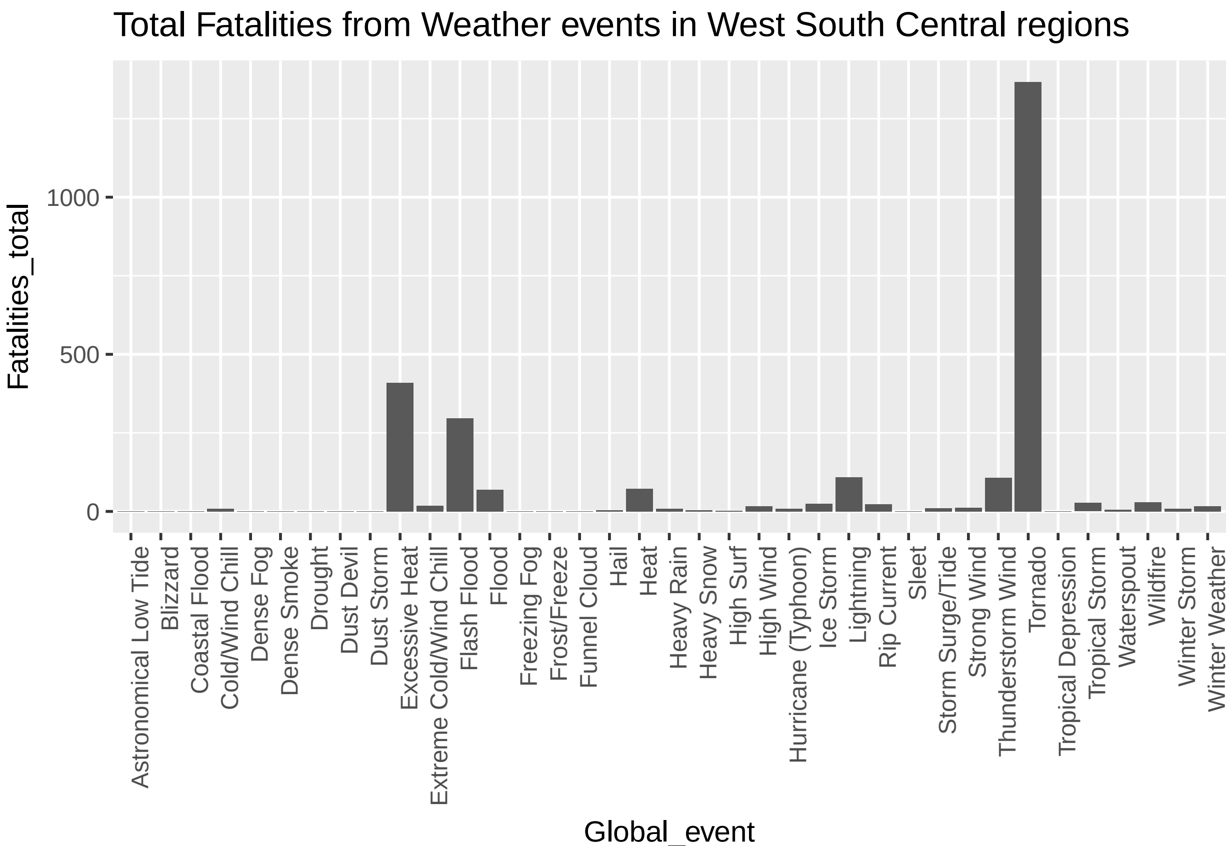



```r
# Tornado
escentral<- ggplot((region_data[grep("East_South_Central", region_data$Region),]),
                   aes(Global_event, Fatalities_total))
escentral + geom_bar(stat='identity')+
  theme(axis.text.x = element_text(angle = 90, hjust = 1))+
  ggtitle('Total Fatalities from Weather events in East South Central regions')
```

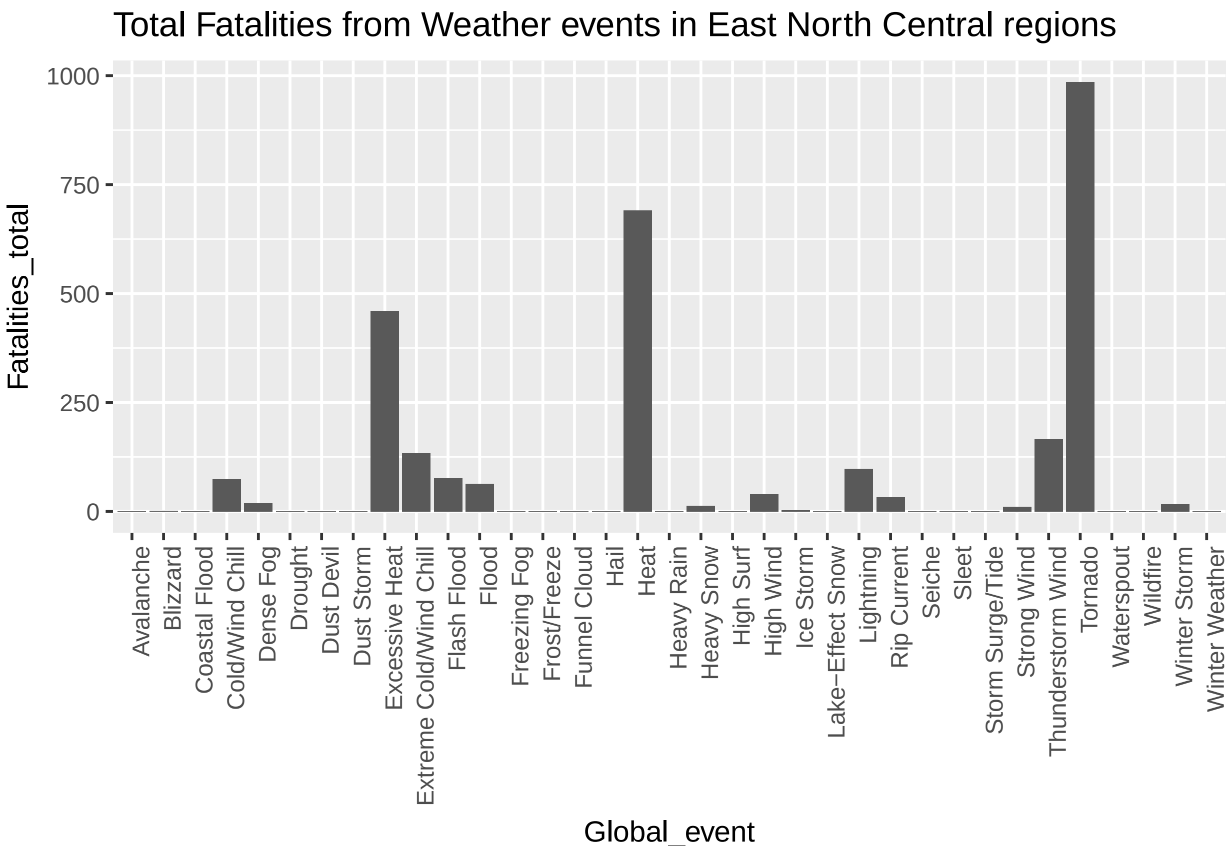


Looking at the above graphs,we can see that within New England, West North Central and East South Central regions of the US, fatalities from weather events are primarily caused by Tornados with few fatalities from other events.

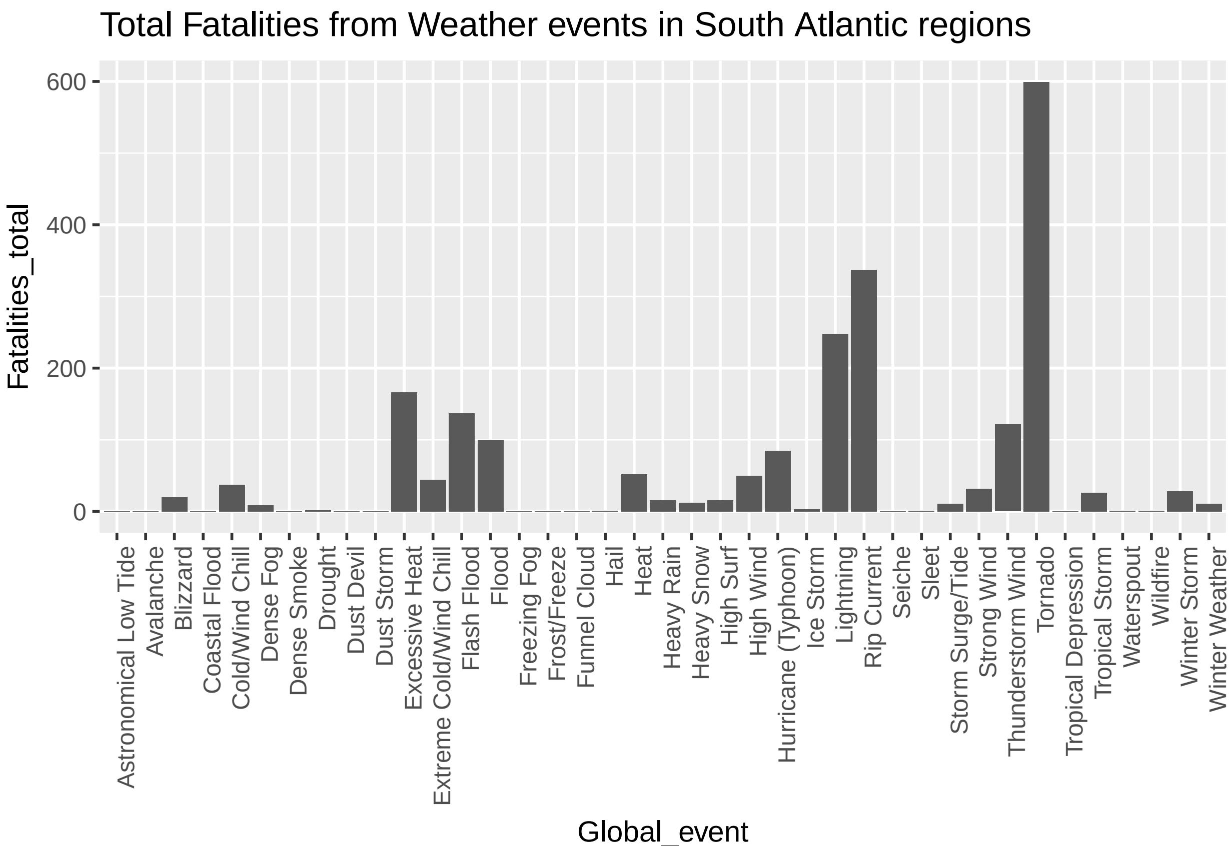
```
# Tornado, heat, flooding
wscentral<- ggplot((region_data[grep("West_South_Central", region_data$Region),]),
                   aes(Global_event, Fatalities_total))
wscentral + geom_bar(stat='identity')+
  theme(axis.text.x = element_text(angle = 90, hjust = 1))+
  ggtitle('Total Fatalities from Weather events in West South Central regions')
```




```r
# Tornado, heat, excessive
encentral<- ggplot((region_data[grep("East_North_Central", region_data$Region),]),
                   aes(Global_event, Fatalities_total))
encentral + geom_bar(stat='identity')+
  theme(axis.text.x = element_text(angle = 90, hjust = 1))+
  ggtitle('Total Fatalities from Weather events in East North Central regions')
```

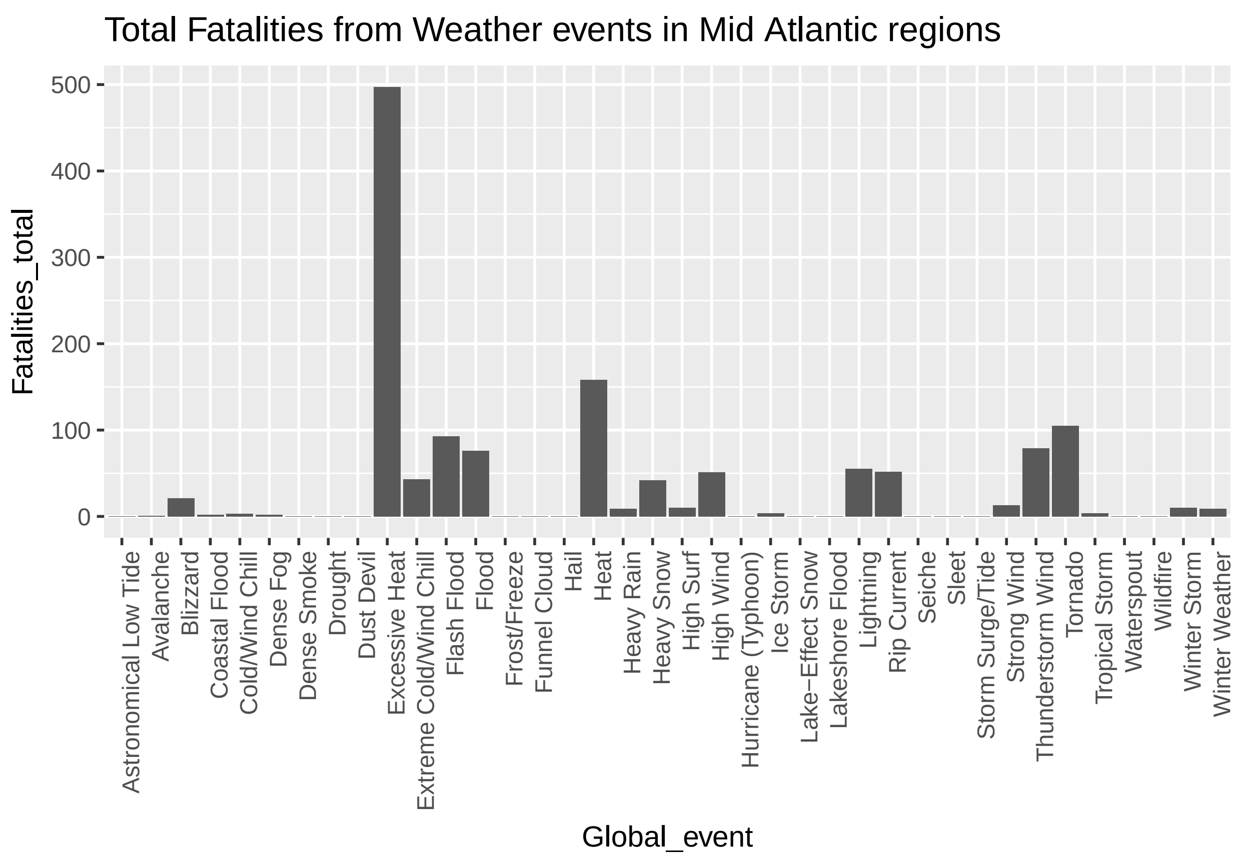



```
# Tornado, rip, lightning
satlantic<- ggplot((region_data[grep("South_Atlantic", region_data$Region),]),
                   aes(Global_event, Fatalities_total))
satlantic + geom_bar(stat='identity')+
  theme(axis.text.x = element_text(angle = 90, hjust = 1))+
  ggtitle('Total Fatalities from Weather events in South Atlantic regions')
```

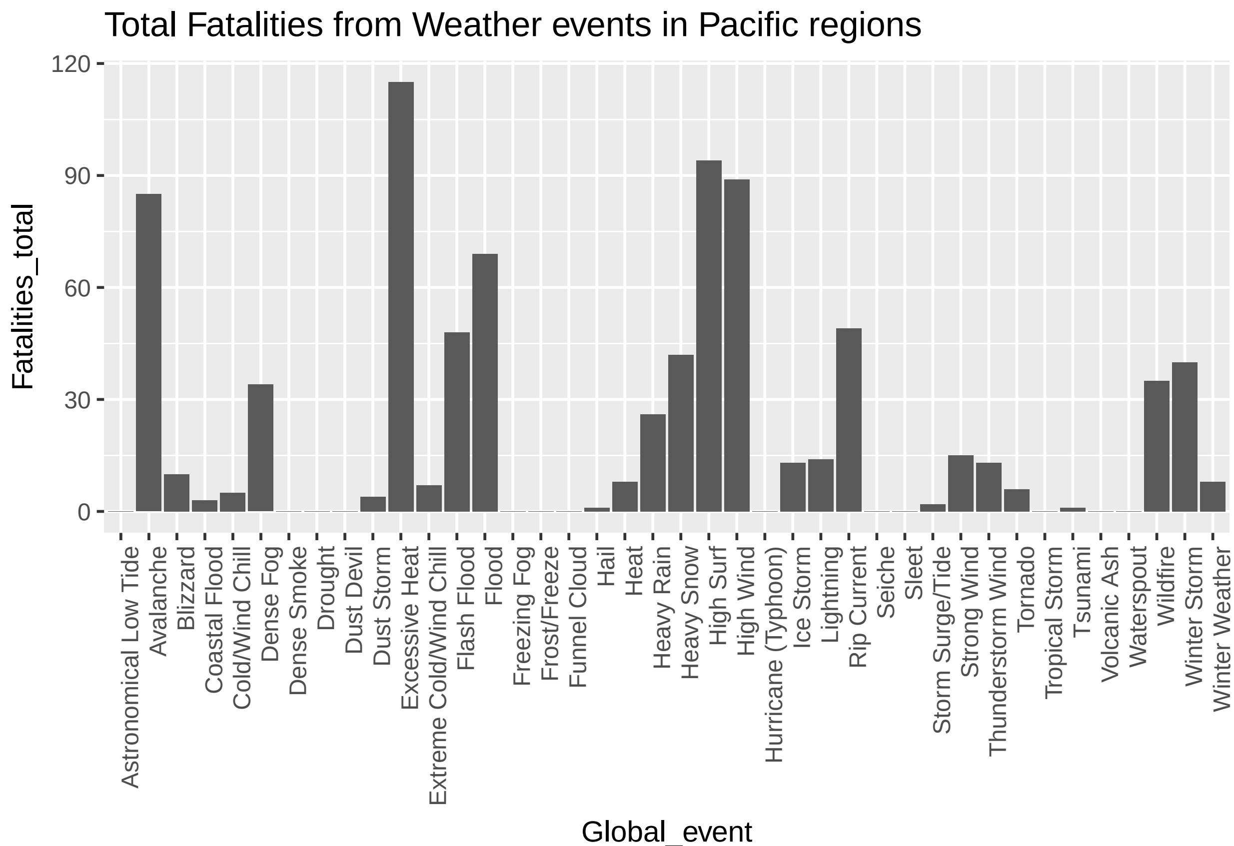


Both West South Central ,East North Central and South Atlantic also showed the most fatalities from Tornados. However, each show high fatalities from Heat, Flooding, Rip Currents or Lightning.

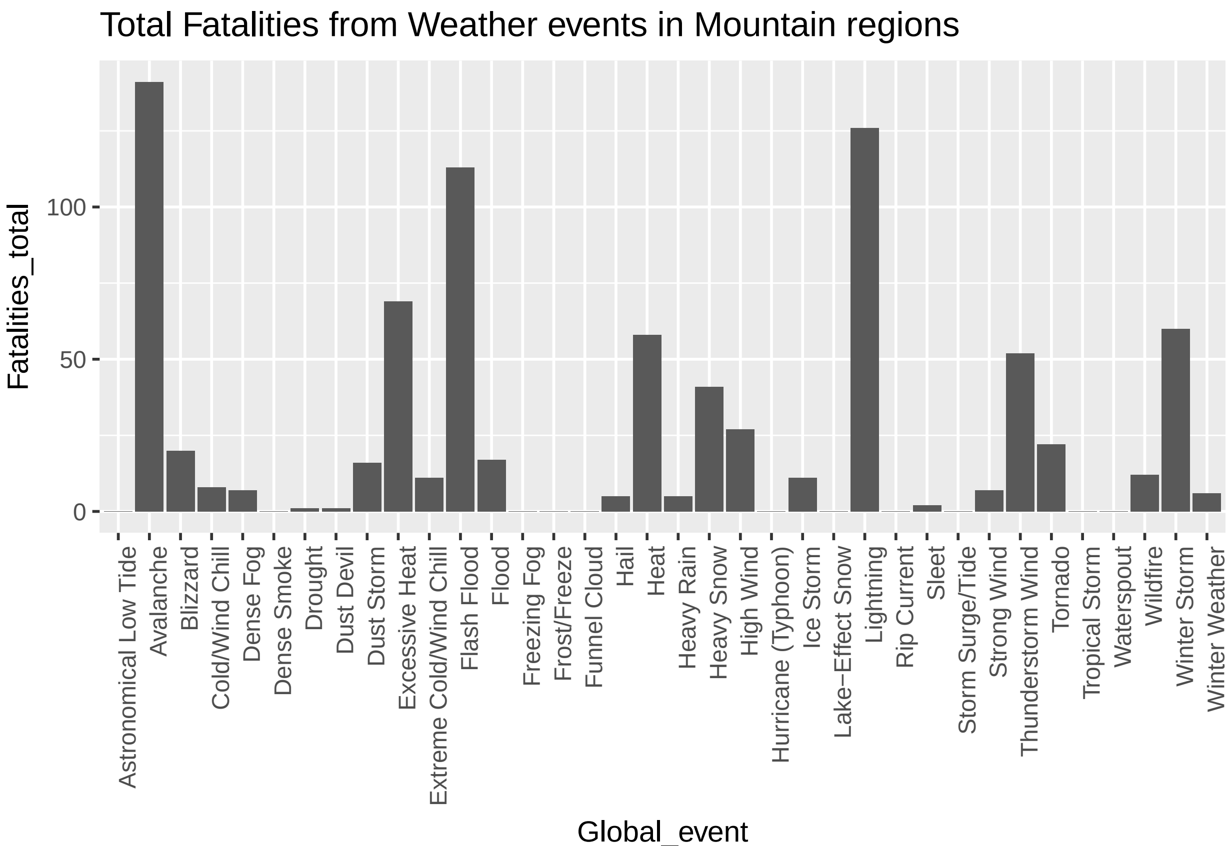
```
#Excessive Heat and Heat
atlantic<- ggplot((region_data[grep("Mid_Atlantic", region_data$Region),]),
                  aes(Global_event, Fatalities_total))
atlantic + geom_bar(stat='identity')+
  theme(axis.text.x = element_text(angle = 90, hjust = 1))+
  ggtitle('Total Fatalities from Weather events in Mid Atlantic regions')
```




```r
# Excessive heat, surf, wind, avalance
pacific<- ggplot((region_data[grep("Pacific", region_data$Region),]),
                 aes(Global_event, Fatalities_total))
pacific + geom_bar(stat='identity')+
  theme(axis.text.x = element_text(angle = 90, hjust = 1))+
  ggtitle('Total Fatalities from Weather events in Pacific regions')
```



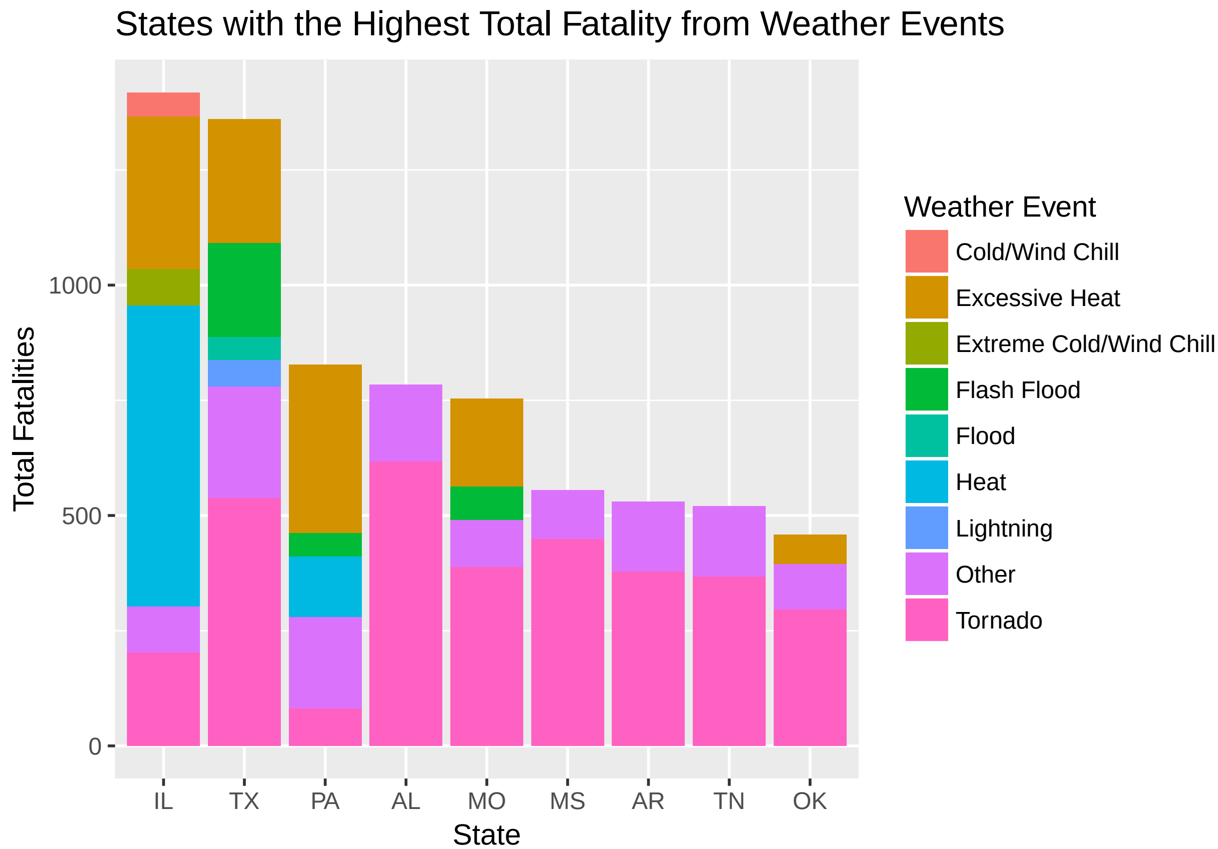

```r
# Avalanche, Flash, Lightning
mountain<- ggplot((region_data[grep("Mountain", region_data$Region),]),
                  aes(Global_event, Fatalities_total))
mountain + geom_bar(stat='identity')+
  theme(axis.text.x = element_text(angle = 90, hjust = 1))+
  ggtitle('Total Fatalities from Weather events in Mountain regions')
```



The final three regions show a more varied picture in terms of the most hazardous weather event. Atlatic regions show a peak fatality rate for Heat events (both Heat and Excessive). Pacific regions simialrly show the highest fatalities from excessive heat, however also exhibit fatalities from Avalances, Flooding, High Winds and Surf. Mountain regions also show a varied influence from weather events with the highest fatalities from Avalanche,Flooding and Lighting.

```r
# This takes a subset for the top 10 states by fatalities
# Removes any NA values
top10states <- as.character(head(state_type_data$STATE, 10))
top10data <- state_type_data[(grep(paste(top10states, collapse ="|"),
                                   state_type_data$STATE)),]
top10data <- filter(top10data, !is.na(Global_event))

# This groups all the events where there were less
# than 50 fatalities for each event type into an  'Other' category
for (i in 1:dim(top10data)[1]){
if (top10data$Fatalities_total[i] <50){
top10data$Global_event[i] <- 'Other'}}

# Plots the top 10 event types by total fatality
#
g <- ggplot(top10data,
            aes(STATE,Fatalities_total,fill=as.factor(Global_event),
                order=Global_event))
g +  geom_bar(stat = 'identity')+
```

```r
  aes(x=reorder(STATE,-Fatalities_total,sum))+
  ggtitle("States with the Highest Total Fatality from Weather Events")+
  xlab("State")+ylab("Total Fatalities")+labs(fill='Weather Event')
```



Similarly, if we look at the 10 states with the highest number of fatality, they are largely states found within eastern and southen areas of the USA. For 6 of these states, tornados show the greatest fatalities. For the remaining 4, IL, PA and CA showed the higest fatalities from Heat or Excessive Heat while FL's fatalities came from Rip Currents. In the case below, any event type with less than 50 fatalities has been grouped togetehr under the 'Other' category

RECOMMENDATIONS: These results show that ,on the whole, more fatalities tend to occur from weather events within eastern regions of the US. Such fatalities also typically tend to be down to **Tornado** weather events meaning resources to predict and prevent should be directed towards these eastern US regions. The most dangerous weather event for the population of the western coast appear to be less predictable, with highest death rates reported from **Avalanches,High winds and Surf, Lightning, Flooding and Heat events**.

Across the United States, which types of events have the greatest economic consequences?

```r
# Total property and crop damage cost for each event type
type_grouped <- group_by(clean_dataset, Global_event)
cost_data <- summarise(type_grouped,Property_total=sum(PROPDMG),Crops_total=sum(CROPDMG))
#Melt data and remove NA rows
```

```
melted_cost <- melt(cost_data, id.vars = "Global_event")
melted_cost <- filter(melted_cost, !is.na(Global_event))

# Plots the top 10 event types by total fatality
g <- ggplot(melted_cost, aes(Global_event,value/1000000, fill=factor(variable)))
g + geom_bar(stat='identity', position='dodge')+
  theme(axis.text.x = element_text(angle = 90, hjust = 1))+
  facet_grid(melted_cost$variable~., scales = 'free')+
  ggtitle('Total Financial costs from Weather events across the US')+
  xlab("Weather Event")+ylab("Total Cost (Million US$)")+
  theme(legend.position="none")

# Calculate the higehest 2 property damage costs
highestcost <-cost_data$Property_total[order(-cost_data$Property_total)[1]]
secondcost <-cost_data$Property_total[order(-cost_data$Property_total)[2]]
# Calculate the higehest 2 crop damage costs
highestcostcrop <-cost_data$Crops_total[order(-cost_data$Crops_total)[1]]
secondcostcrop <-cost_data$Crops_total[order(-cost_data$Crops_total)[2]]
```

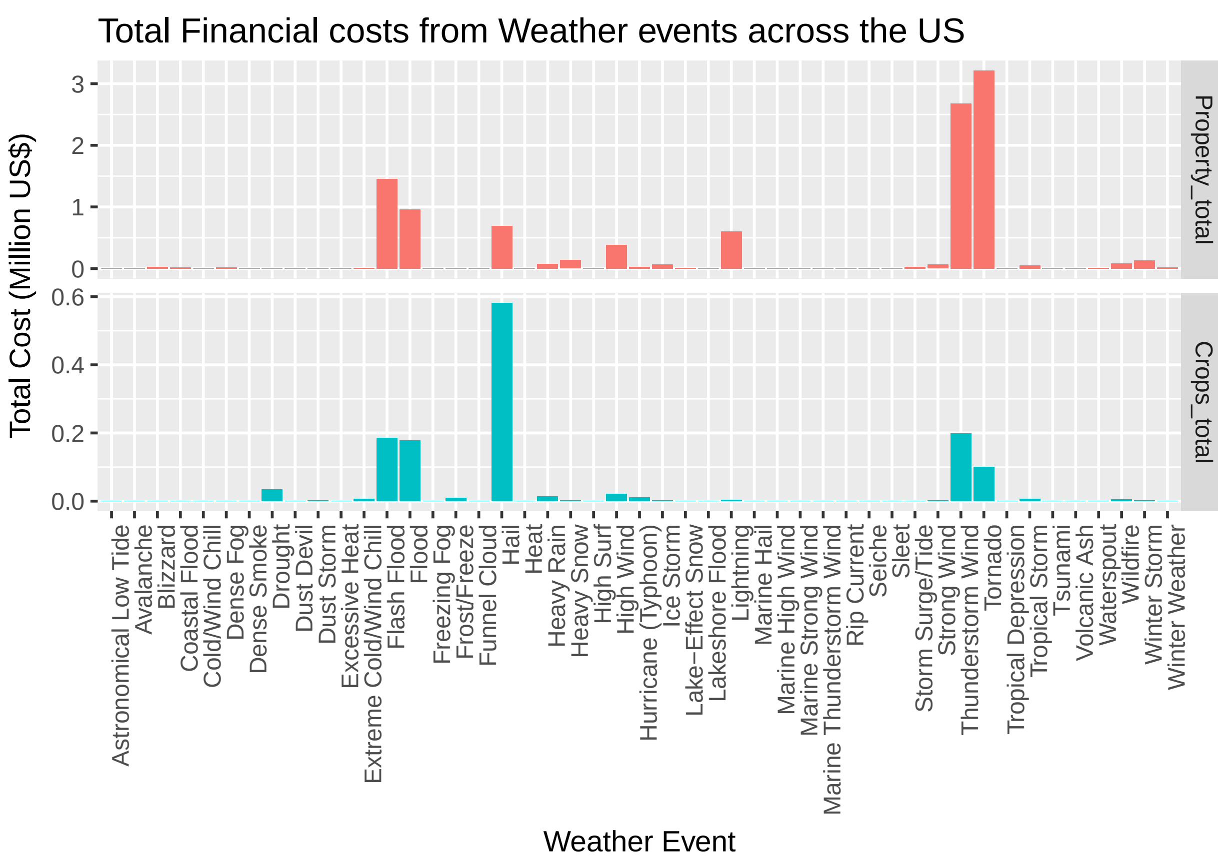


Figure 3: Figure 3: Bar graph to show the total damage cost to both property and crops from each weather event.

In terms of Property damage, there has been 0.011 billion USD worth of damage compared to 0.001 billion USD for crop damage.

Property damage has been largely caused by **Floods** and **Hurricanes (Typhoon)**, with 0.003 Billion USD and 0.003 billion USD respectivly. This was then followed by **Tornados** and **Storm Surges and Tides**. In terms of crop damage however, this was unsuprisingly largely due to **Droughts** and **Floods**, causing 0.001 bilion USD and 0 billion USD worth of damge respectivly.

Appendix

Code used to clean the original dataset and to categorise the weatehr events into the correct groups.

```
# Some inputs have not been given the standard EV Type name
# so we want to ensure all the same event type, regardless of
# exact naming convention, are analysed under the same
# category.  Create a new coolumn variable to save his
# category information
clean_dataset[, 38] <- NA
colnames(clean_dataset)[38] <- "Global_event"
# Make all the EVTYPE enries lower case to avoid duplciates
clean_dataset$EVTYPE <- tolower(clean_dataset$EVTYPE)

# Create a variable to store the inputted event names
# associated with each of the 48 weather event types For
# example 'Heavy Rain', 'Heavy Precipitation' and 'Torrential
# Rain' will all be categorised as 'Heavy Rain'
event_types <- vector(mode = "list", length = 48)
# Create a variable to save the index locations of entries
# associated with each event type
event_index <- vector(mode = "list", length = 48)

# Predefined names of the 48 weather event types
evnames <- c("Astronomical Low Tide", "Avalanche", "Blizzard",
    "Coastal Flood", "Cold/Wind Chill", "Debris Flow", "Dense Fog",
    "Dense Smoke", "Drought", "Dust Devil", "Dust Storm", "Excessive Heat",
    "Extreme Cold/Wind Chill", "Flash Flood", "Flood", "Frost/Freeze",
    "Funnel Cloud", "Freezing Fog", "Hail", "Heat", "Heavy Rain",
    "Heavy Snow", "High Surf", "High Wind", "Hurricane (Typhoon)",
    "Ice Storm", "Lake-Effect Snow", "Lakeshore Flood", "Lightning",
    "Marine Hail", "Marine High Wind", "Marine Strong Wind",
    "Marine Thunderstorm Wind", "Rip Current", "Seiche", "Sleet",
    "Storm Surge/Tide", "Strong Wind", "Thunderstorm Wind", "Tornado",
    "Tropical Depression", "Tropical Storm", "Tsunami", "Volcanic Ash",
    "Waterspout", "Wildfire", "Winter Storm", "Winter Weather")
names(event_types) <- evnames
names(event_index) <- evnames

# Creates a list of all the unique inputs for the evtype
# variable
labels <- unique(as.character(clean_dataset$EVTYPE))

# This chunk of text goes through each of the 48 variables
# and and finds the associated EVTYPE variable names and the
# indices of the entries in the main dataset.  Each of these
```

```r
# were manually checked to ensure each search criteria
# encompased all the relavant entries In some cases, when the
# event type still stated 2 categories, the first category
# inputted was chosen.  Similarly, some evtype inputs did not
# fit into any category or were not specific enough to be
# assogned to 1 cateogy or the other e.g. 'Record high' may
# refer to heat, rain or wind, so was not included.
event_index[["High Wind"]] <- grep("(?=.*wnd.*)(?!.*tstm)|(?=.*wind.*)(?!.*low)(?!.*marine)(?!.*strong)
    x = clean_dataset$EVTYPE, ignore.case = TRUE, perl = TRUE)
event_index[["Tornado"]] <- grep("(?=.*torn.*)(?!.*waterspout)",
    x = clean_dataset$EVTYPE, ignore.case = TRUE, perl = TRUE)
event_index[["Dust Devil"]] <- grep("devil|devel", x = clean_dataset$EVTYPE,
    ignore.case = TRUE)
event_index[["Coastal Flood"]] <- grep("(?=.*coastal.*)(?!.*surf)(?!.*erosion)|(?=.*cstl.*)|(?=.*tidal*)
    x = clean_dataset$EVTYPE, ignore.case = TRUE, perl = TRUE)
event_index[["Lightning"]] <- grep("(?=.*^lig.*ing)(?!.*rain)|(?=.*^ lig.*ing)",
    x = clean_dataset$EVTYPE, ignore.case = TRUE, perl = TRUE)
event_index[["Lakeshore Flood"]] <- grep(c("Lake.*flood"), x = clean_dataset$EVTYPE,
    ignore.case = TRUE)
event_index[["Winter Storm"]] <- grep("^win.*sto", x = clean_dataset$EVTYPE,
    ignore.case = TRUE)
event_index[["Winter Weather"]] <- grep("^win.*wea", x = clean_dataset$EVTYPE,
    ignore.case = TRUE)
event_index[["Seiche"]] <- grep("Seiche", x = clean_dataset$EVTYPE,
    ignore.case = TRUE)
event_index[["Wildfire"]] <- grep("(?=.*fire)(?!.*lightning)",
    x = clean_dataset$EVTYPE, ignore.case = TRUE, perl = TRUE)
event_index[["Waterspout"]] <- grep("^Wa.*ter.*spout|^ Wa.*ter.*spout",
    x = clean_dataset$EVTYPE, ignore.case = TRUE)
event_index[["Hurricane (Typhoon)"]] <- grep("Hurricane|Typhoon",
    x = clean_dataset$EVTYPE, ignore.case = TRUE)
event_index[["Thunderstorm Wind"]] <- grep("^(?=.*TSTM)(?!.*tornado)(?!.*marine)(?!.*non)|^(?=.*T.*e.*o
    x = clean_dataset$EVTYPE, ignore.case = TRUE, perl = TRUE)
event_index[["Tropical Storm"]] <- grep("^Tropical Storm", x = clean_dataset$EVTYPE,
    ignore.case = TRUE)
event_index[["Tropical Depression"]] <- grep("depression", x = clean_dataset$EVTYPE,
    ignore.case = TRUE)
event_index[["Flash Flood"]] <- grep("Flash", x = clean_dataset$EVTYPE,
    ignore.case = TRUE)
event_index[["Hail"]] <- grep("^(?=.*hail)(?!.*marine)(?!.*wind)",
    x = clean_dataset$EVTYPE, ignore.case = TRUE, perl = TRUE)
event_index[["Heavy Rain"]] <- grep("^(?=.*Precip)(?!.*monthly)|(?=.*rain)(?!.*monthly)(?!.*freezing)(?
    x = clean_dataset$EVTYPE, ignore.case = TRUE, perl = TRUE)
event_index[["Drought"]] <- grep("drought", x = clean_dataset$EVTYPE,
    ignore.case = TRUE)
event_index[["Funnel Cloud"]] <- grep("(?=.*funnel)(?!.*thunderstorm)(?!.*waterspout)",
    x = clean_dataset$EVTYPE, ignore.case = TRUE, perl = TRUE)
event_index[["Frost/Freeze"]] <- grep("(?=.*frost)|(?=.*freeze)(?!.*snow)",
    x = clean_dataset$EVTYPE, ignore.case = TRUE, perl = TRUE)
event_index[["Lake-Effect Snow"]] <- grep("lake.* snow", x = clean_dataset$EVTYPE,
    ignore.case = TRUE)
event_index[["Heavy Snow"]] <- grep("(?=.*snow)(?!.*lake)(?!.*lake)",
    x = clean_dataset$EVTYPE, ignore.case = TRUE, perl = TRUE)
```

```r
event_index[["Ice Storm"]] <- grep("(?=.*ice.*storm)|(?=.*freezing)(?!.*fog)(?!.*snow)(?!.*road)",
    x = clean_dataset$EVTYPE, ignore.case = TRUE, perl = TRUE)
event_index[["Heat"]] <- grep("(?=.*heat)(?!.*ex.*)|(?=.*warm)|(?=.*hot)|(?=.*HIGH.*TEMP)",
    x = clean_dataset$EVTYPE, ignore.case = TRUE, perl = TRUE)
event_index[["Excessive Heat"]] <- grep("(?=.*ex.*heat)", x = clean_dataset$EVTYPE,
    ignore.case = TRUE, perl = TRUE)
event_index[["Volcanic Ash"]] <- grep("volcanic", x = clean_dataset$EVTYPE,
    ignore.case = TRUE)
event_index[["Flood"]] <- grep("(?=.*flood)(?!.*urban)(?!.*flash)(?!.*street)(?!.*urban)(?!.*tidal)(?!.
    x = clean_dataset$EVTYPE, ignore.case = TRUE, perl = TRUE)
event_index[["Marine Hail"]] <- grep("marine.*hail", x = clean_dataset$EVTYPE,
    ignore.case = TRUE)
event_index[["Marine High Wind"]] <- grep("marine.*high", x = clean_dataset$EVTYPE,
    ignore.case = TRUE)
event_index[["Marine Strong Wind"]] <- grep("marine.*strong",
    x = clean_dataset$EVTYPE, ignore.case = TRUE)
event_index[["Marine Thunderstorm Wind"]] <- grep("marine.*tstm|marine.*thunder",
    x = clean_dataset$EVTYPE, ignore.case = TRUE)
event_index[["Extreme Cold/Wind Chill"]] <- grep("(?=.*ext.*chill)|(?=.*ex.*cold)|(?=.*sev.*cold)",
    x = clean_dataset$EVTYPE, ignore.case = TRUE, perl = TRUE)
event_index[["Cold/Wind Chill"]] <- grep("(?=.*wind.*chill)(?!.*ex.*)(?!.*snow)|(?=.*cold)(?!.*ex.*)(?!
    x = clean_dataset$EVTYPE, ignore.case = TRUE, perl = TRUE)
event_index[["Dense Fog"]] <- grep("(?=.*fog)(?!.*ice)(?!.*freezing)",
    x = clean_dataset$EVTYPE, ignore.case = TRUE, perl = TRUE)
event_index[["Freezing Fog"]] <- grep("(?=.*fog)(?=.*freezing)|(?=.*fog)(?=.*ice)",
    x = clean_dataset$EVTYPE, ignore.case = TRUE, perl = TRUE)
event_index[["Sleet"]] <- grep("(?=.*sleet)(?!.*snow)", x = clean_dataset$EVTYPE,
    ignore.case = TRUE, perl = TRUE)
event_index[["Rip Current"]] <- grep("rip", x = clean_dataset$EVTYPE,
    ignore.case = TRUE)
event_index[["High Surf"]] <- grep("(?=.*surf)(?!.*rip)|(?=.*waves)(?!.*heat)|(?=.*swell)",
    x = clean_dataset$EVTYPE, ignore.case = TRUE, perl = TRUE)
event_index[["Astronomical Low Tide"]] <- grep("ASTRONOMICAL LOW TIDE",
    x = clean_dataset$EVTYPE, ignore.case = TRUE)
event_index[["Avalanche"]] <- grep("Aval.*|rock|.*slide", x = clean_dataset$EVTYPE,
    ignore.case = TRUE)
event_index[["Dense Smoke"]] <- grep("smoke", x = clean_dataset$EVTYPE,
    ignore.case = TRUE)
event_index[["Tsunami"]] <- grep("TSUNAMI", x = clean_dataset$EVTYPE,
    ignore.case = TRUE)
event_index[["Blizzard"]] <- grep("BLIZZARD", x = clean_dataset$EVTYPE,
    ignore.case = TRUE)
event_index[["Dust Storm"]] <- grep("(?=.*dust)(?!.*dev.*l)(?!.*wind)",
    x = clean_dataset$EVTYPE, ignore.case = TRUE, perl = TRUE)
event_index[["Storm Surge/Tide"]] <- grep("storm surge|high tides",
    x = clean_dataset$EVTYPE, ignore.case = TRUE)
event_index[["Strong Wind"]] <- grep("(?=.*strong.*wind.*)(?!.*marine)",
    x = clean_dataset$EVTYPE, ignore.case = TRUE, perl = TRUE)

# Here we loop though each of the 48 weather conditions and
# save out the associated evntype names and we alter the
# Global_type variable to represent the associated global
# event category (out of a possibel 48)
```

```r
used_labels <- c()
for (i in (1:(length(event_index)))) {
    this_name <- names(event_types)[i]
    clean_dataset$Global_event[event_index[[this_name]]] <- this_name
    used_labels <- c(used_labels, (event_types[[this_name]]))
}

# Here we save out the unused names for the EVTYPE variable -
# this was again checked to ensure none met the desires
# criteria
nonlabels <- labels[-(grep(paste("!", paste(used_labels, collapse = "|",
    sep = "|")), labels, ignore.case = TRUE))]
```