Impact of Severe Weather Events on Public Health and Economy in the US

Lisa Yuan

February 25, 2016

Synonpsis

The purpose of this report is to show the harmful impact of severe weather events on both public health and the economy. The analysis is performed based on the storm database collected from the U.S. National Oceanic and Atmospheric Administration (NOAA) from 1950-2011. The report identifies the weather event types that has the most significant impact on population health (as measured by fatalities and injuries) and on the economy (as measured by total damage to property and crops combined in current US dollars).

Based an my analysis, I find that the severe weather events associated with wind, storm, heat and drought cause the most damage to public health, whereas severe weather events associated with flood cause the most damage to the economy.

Working Directory and Libraries

```r
setwd("C:/Users/Lisa/version-control/RepData_PeerAssessment2")
library(R.utils)
```

```
## Loading required package: R.oo
```

```
## Loading required package: R.methodsS3
```

```
## R.methodsS3 v1.7.1 (2016-02-15) successfully loaded. See ?R.methodsS3 for help.
```

```
## R.oo v1.20.0 (2016-02-17) successfully loaded. See ?R.oo for help.
```

```
##
## Attaching package: 'R.oo'
```

```
## The following objects are masked from 'package:methods':
##
##     getClasses, getMethods
```

```
## The following objects are masked from 'package:base':
##
##     attach, detach, gc, load, save
```

```
## R.utils v2.2.0 (2015-12-09) successfully loaded. See ?R.utils for help.
```

```
##
## Attaching package: 'R.utils'
```

```
## The following object is masked from 'package:utils':
##
##     timestamp

## The following objects are masked from 'package:base':
##
##     cat, commandArgs, getOption, inherits, isOpen, parse, warnings
```

```
library(ggplot2)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(grid)
library(gridExtra)
```

Data Processing

First, I created a folder for the storm data to be stored. Then, I downloaded and unzipped the file.

```
## Create the directory data where the storm data will be downloaded into
if(!file.exists("./data")){
        dir.create("./data")
}

# Download file
if (!"stormData.csv.bz2" %in% dir("./data/")) {
        print ("lets download file")
        download.file("https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2",
                      destfile="data/stormData.csv.bz2")
}

#unzip bz2 file to csv file
if(!file.exists ("./data/StormData.csv")) {
        filePath <- "./data/StormData.csv.bz2"
        destPath <- "./data/StormData.csv"
        bunzip2(filePath, destPath, overwrite=TRUE, remove=FALSE)
}
```

Then, I read the csv file into R environment.

```r
if(!"stormData" %in% ls()) {
stormData <- read.csv("./Data/StormData.csv", header = TRUE, sep = ",")
}
```

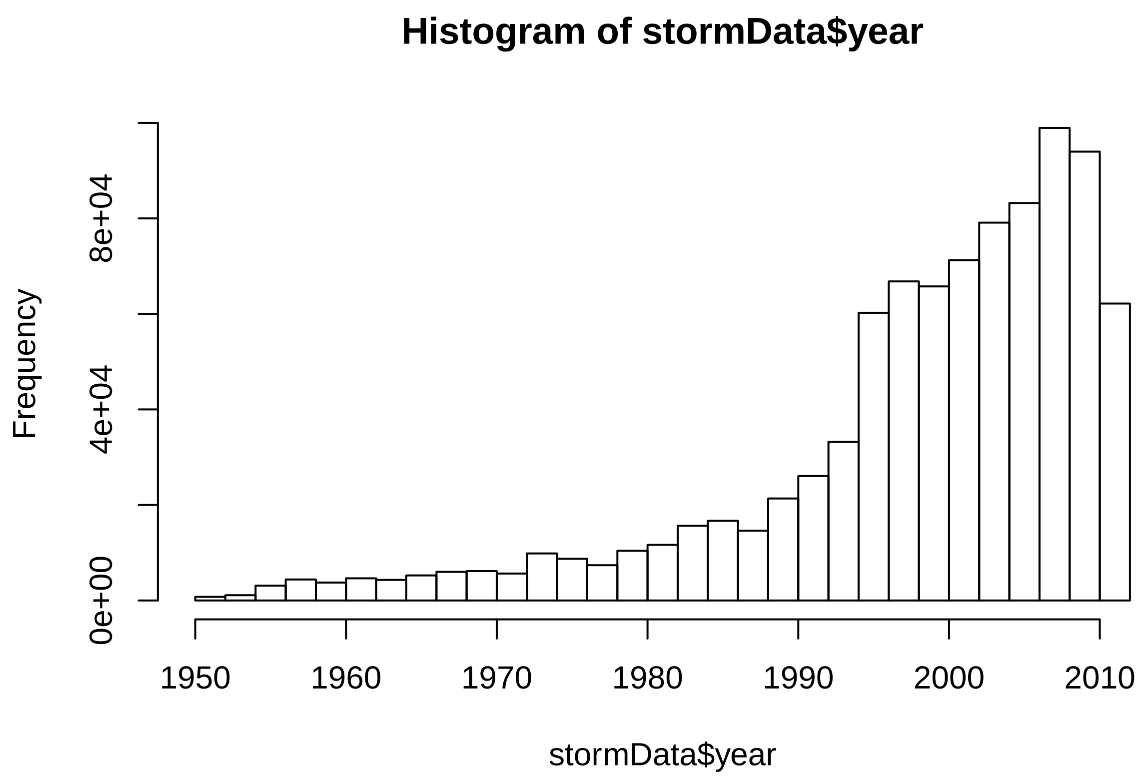
After the stormData is created as a dataframe, we're able to examine the data. It contains 902,297 rows and 38 columns. The dataset covers the times span from 1950 to 2011. With a quick histogram, it is clearly evident that the earlier years in the dataset have fewer recorded events, and starting 1994, we see significant increase of the amount of data available.

```r
# Convert the BGN_DATE to year

stormData$year <- as.numeric(format(as.Date(stormData$BGN_DATE, format = "%m/%d/%Y %H:%M:%S"), "%Y"))

# Visulization with histogram

hist(stormData$year, breaks=30)
```



Based on the information from the above histogram and the purpose of this research, I subset data keeping only useful columns and rows with year starting at 1990.

```r
#Reduce the datasize, only keep the necessary columes
reduced_stormData<- select(stormData, STATE, EVTYPE, FATALITIES, INJURIES, PROPDMG,
                           PROPDMGEXP, CROPDMG, CROPDMGEXP, year)
```

By examing the EVTYPE, I have found that the event types can be consolidated into 12 catagories. These catagories are:

- Precipitation and Fog
- Dust
- Fire and Volcano
- Flood
- Heat and Drought
- Landslide and Erosion
- Ocean Weather
- Precipitation and Fog
- Thunderstorm and Lightning
- Wind and Storm
- Wintery Weather
- Other

```
#Consolidate the EVTTYPE and created a new variable called EVTCatagory

reduced_stormData$EVTCatagory <- NA
reduced_stormData[grepl("precipitation|rain|hail|drizzle|wet|percip|burst|depression|
                         fog|wall cloud", reduced_stormData$EVTYPE, ignore.case=TRUE),
                   "EVTCatagory"] <- "Precipitation and Fog"
reduced_stormData[grepl("wind|storm|wnd|tornado|waterspout", reduced_stormData$EVTYPE,
                         ignore.case = TRUE), "EVTCatagory"] <- "Wind & Storm"
reduced_stormData[grepl("Slide|erosion|slump", reduced_stormData$EVTYPE, ignore.case=TRUE),
                   "EVTCatagory"] <- "Landslide and Erosion"
reduced_stormData[grepl("warmth|warm|heat|dry|hot|drought|thermia|temperature record|
                         record temperature|record high", reduced_stormData$EVTYPE,
                         ignore.case = TRUE), "EVTCatagory"] <- "Heat and Drought"
reduced_stormData[grepl("cold|chill|ice|Wind Chill|Frost|Freeze|Winter Weater|Winter|
                        Blizzard|Snow|Avalanche", reduced_stormData$EVTYPE, ignore.case=TRUE),
                   "EVTCatagory"] <- "Wintery Weather"
reduced_stormData[grepl("Tide|Marine|Current|Tsunami|wave|seas|surf|typhoon|hurricane", reduced_stormDa
                        ignore.case=TRUE), "EVTCatagory"] <- "Ocean Weather"
reduced_stormData[grepl("flood|blow-out|swells|fld|dam break", reduced_stormData$EVTYPE,
                         ignore.case=TRUE), "EVTCatagory"] <- "Flood"
reduced_stormData[grepl("dust storm|dust devil", reduced_stormData$EVTYPE, ignore.case=TRUE),
                   "EVTCatagory"] <- "Dust"
reduced_stormData[grepl("fire|volcanic", reduced_stormData$EVTYPE, ignore.case=TRUE),
                   "EVTCatagory"] <- "Fire and Volcano"
reduced_stormData[grepl("Thunderstorm|lightning", reduced_stormData$EVTYPE, ignore.case=TRUE),
                   "EVTCatagory"] <- "Thunderstorm and Lightning"
reduced_stormData$EVTCatagory[is.na(reduced_stormData$EVTCatagory)] <- "Other"
```

I cast the EVTCatagory to a factor variable for later ploting and removed the EVTYPE column.

```
reduced_stormData$EVTCatagory <- as.factor(reduced_stormData$EVTCatagory)
reduced_stormData$EVTYPE <- NULL
```

In order to assess economic damage casued by severe weather, I calculated total amount using helper functions to convert PROPDMG, PROPDMGEXP into PropDamage, and CROPDMG, CROPDMGEXP into CropDamage

```
# Calculate the damage with conversion
```

```r
expConvert <- function(x) {
        if(is.numeric(x)) {
                x <- x
        }
        else if (grepl("h", x, ignore.case=TRUE)) {
                x <- 2
        }
        else if (grepl("k", x, ignore.case=TRUE)) {
                x <- 3
        }
        else if (grepl("m", x, ignore.case=TRUE)) {
                x <- 6
        }
        else if (grepl("b", x, ignore.case=TRUE)) {
                x <- 9
        }
        else if(x==""||x==" ") {
                x <- 0
        }
        else{
                x <- NA
        }
        x
}


calamt <- function(num, exp) {
        pow<-expConvert(exp)
        if(is.numeric(num)){
          num <- num * (10^pow)
        }

        if(!is.numeric(num)) {
                num <- 0
        }

        num

}

# Getting the absolute value for PropDamage and CropDamage, then add them together to
# create total damage, 3 new columns are created.

reduced_stormData$propDamage <- mapply(calamt, reduced_stormData$PROPDMG, reduced_stormData$PROPDMGEXP)
reduced_stormData$cropDamage <- mapply(calamt, reduced_stormData$CROPDMG, reduced_stormData$CROPDMGEXP)
reduced_stormData$totalDamage <- reduced_stormData$propDamage + reduced_stormData$cropDamage
```

I removed the columns no longer needed and subset the data so that I only look at the data starting from 1990.

```r
# Remove columns not needed to reduce dataframe size
reduced_stormData$PROPDMG <- NULL
reduced_stormData$PROPDMGEXP <- NULL
```

```
reduced_stormData$CROPDMG <- NULL
reduced_stormData$CROPDMGEXP <- NULL

#Furthur reduce the data set by only track the datapoints after 1990

recent_stormData<- filter(reduced_stormData, year >= 1990)
```

Results

In this section, I aggregated the total number of fatilities and injuries caused by different types of severe weather catagoy.

```
#Aggregate the data by event and sorted by fatility/injuries in the descending order

fatality <- aggregate(FATALITIES ~ EVTCatagory, data=recent_stormData, FUN=sum)
fatality <- fatality %>% arrange(desc(FATALITIES))
injuries <- aggregate(INJURIES ~ EVTCatagory, data=recent_stormData, FUN=sum)
injuries <- injuries %>% arrange(desc(INJURIES))
```

```
#Merge fatality and injuries into one dataset
humanDamage <- merge(fatality, injuries, by="EVTCatagory")
```

Impact on Public Health

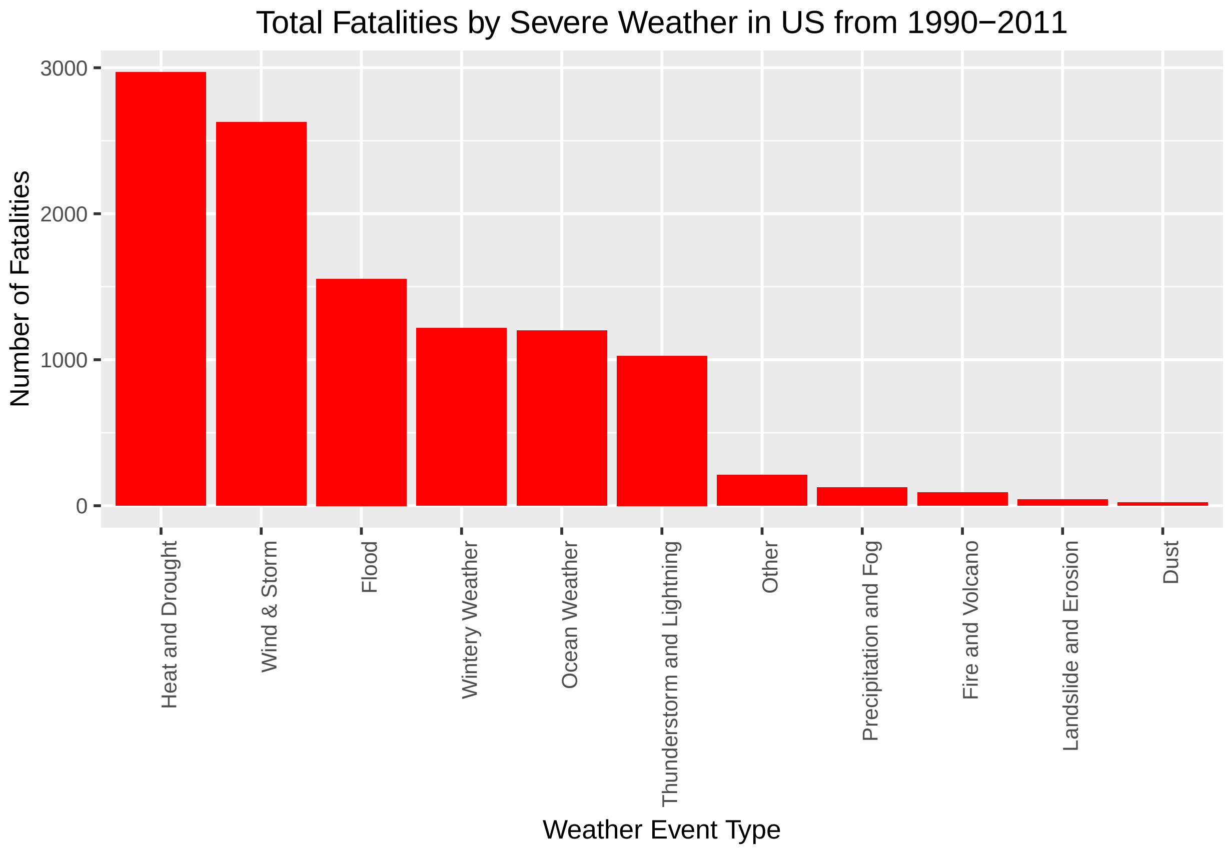
Now, let's examine the severe weather damage to public health caused by different events.

```
humanDamage
```

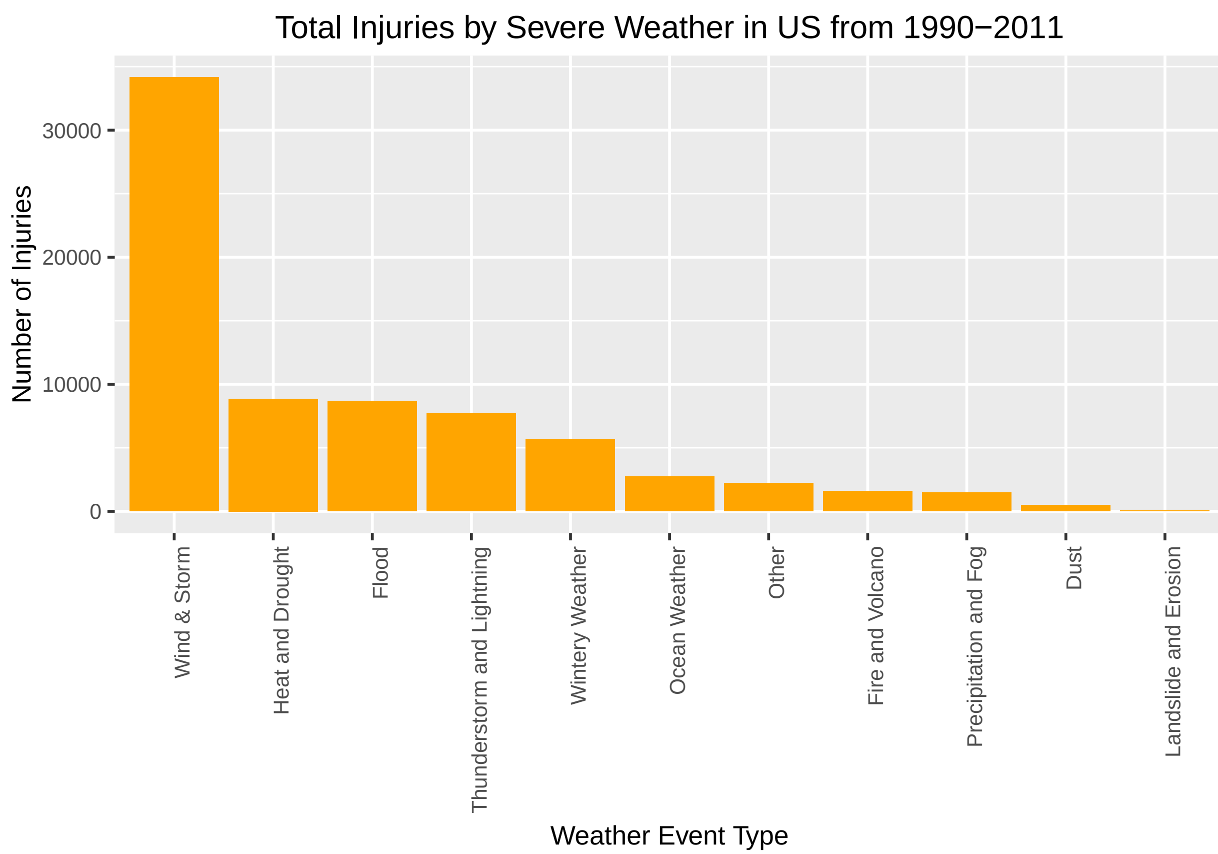
```
##                  EVTCatagory FATALITIES INJURIES
## 1                       Dust         24      483
## 2           Fire and Volcano         90     1608
## 3                      Flood       1554     8685
## 4           Heat and Drought       2969     8865
## 5      Landslide and Erosion         44       55
## 6              Ocean Weather       1199     2724
## 7                      Other        212     2239
## 8      Precipitation and Fog        124     1467
## 9  Thunderstorm and Lightning       1027     7710
## 10               Wind & Storm       2626    34156
## 11           Wintery Weather       1218     5707
```

Below are plots for injuries and fatalities impacted by major weather events ordered by severity.

```
#Fatility plot
ggplot(data=humanDamage, aes(y=FATALITIES, x=reorder(EVTCatagory, -FATALITIES))) +
        geom_bar(stat="identity", fill="red") +
        scale_y_continuous("Number of Fatalities") +
        ggtitle("Total Fatalities by Severe Weather in US from 1990-2011") +
        xlab("Weather Event Type") +
        theme(axis.text.x=element_text(angle=90, hjust=1), text=element_text(size=10))
```




```
#Injury plot
ggplot(data=humanDamage, aes(y=INJURIES, x=reorder(EVTCatagory,-INJURIES))) +
        geom_bar(stat="identity", fill="orange") +
        scale_y_continuous("Number of Injuries") +
        ggtitle("Total Injuries by Severe Weather in US from 1990-2011") +
        xlab("Weather Event Type") +
        theme(axis.text.x=element_text(angle=90, hjust=1), text=element_text(size=10))
```



Based on the above two histogrms, we find that "Wind and Storm" caused most human injuries, and "Heat and Drought" along with "Wind and Storm" contributed to most human fatalities in the US from 1990 to 2011.

Impact on Economy

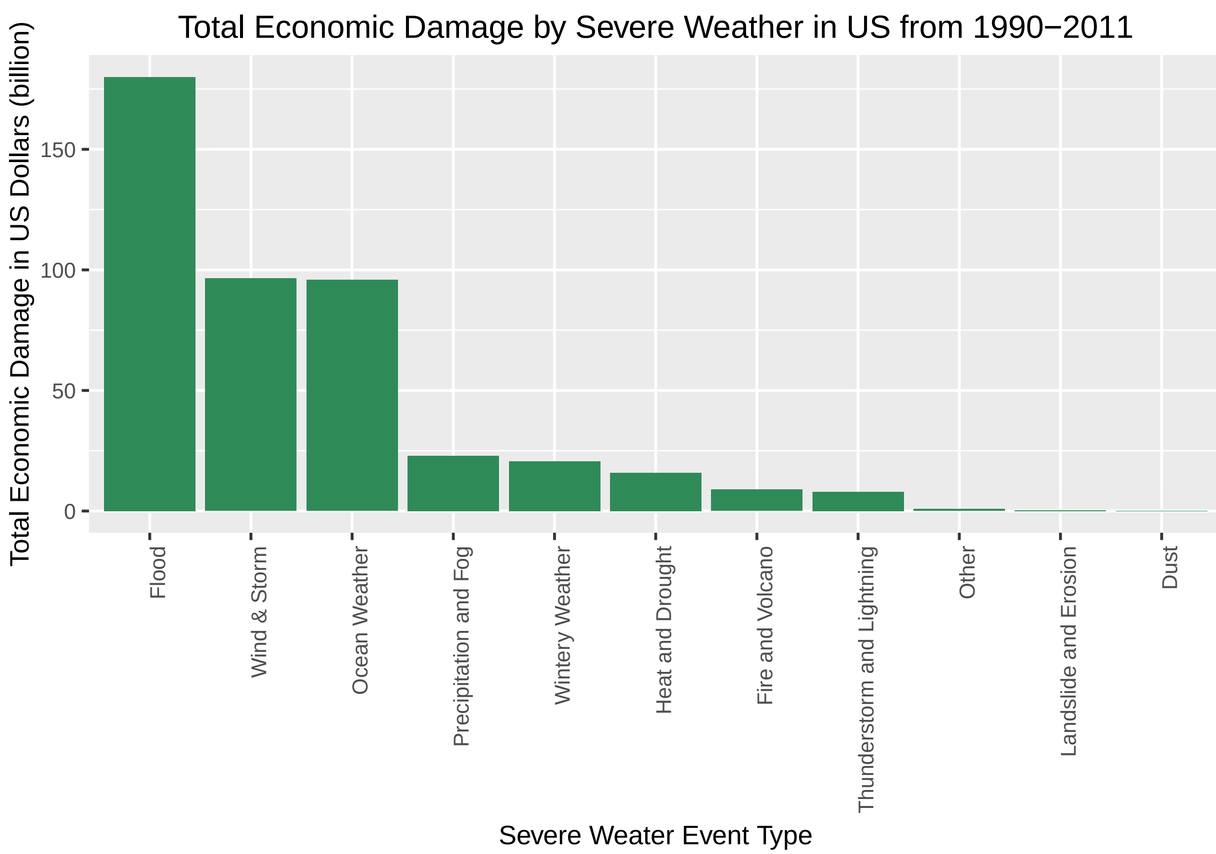
Now, let's examine what types of severe weather caused most damages to economy.

```r
# Aggregate data for total damage
EconDamage <- aggregate(totalDamage~EVTCatagory, data=recent_stormData, FUN=sum)
EconDamage <- EconDamage %>% arrange(desc(totalDamage))
EconDamage
```

```
##                  EVTCatagory  totalDamage
## 1                      Flood 179975699394
## 2               Wind & Storm  96450409816
## 3              Ocean Weather  95850254200
## 4      Precipitation and Fog  22920299720
## 5            Wintery Weather  20639077510
## 6           Heat and Drought  15933954580
## 7           Fire and Volcano   8900410130
## 8 Thunderstorm and Lightning   8036351698
## 9                      Other    802675300
## 10     Landslide and Erosion    348279100
## 11                      Dust      9918130
```

Below is the histogram of total property damage combined with total crop damage affected by severe weather events.

```
ggplot(data=EconDamage, aes(x=reorder(EVTCatagory, -totalDamage), y=(totalDamage/10^9))) +
        geom_bar(stat="identity", fill="seagreen4") +
        scale_y_continuous("Total Economic Damage in US Dollars (billion)") +
        xlab("Severe Weater Event Type") +
        ggtitle("Total Economic Damage by Severe Weather in US from 1990-2011")+
        theme(axis.text.x=element_text(angle=90, hjust=1), text=element_text(size=10))
```



Based on the above histogrm, we can conclude that the "Flood" along wtih "Wind & Storm" and "Ocean Weather" caused the most damage to economy in the United States from 1990 to 2011.

Conclusion

Based on the data collected from the NOAA and my analysis, we can conclude that the most damaging severe weather types are "Wind and Storm" and "Excessive heat and Drought" to public health. Whereas the "Flood", "Wind and Storm" and "Ocean Weather" has the most damaging economic impact.