Reproducible Research: Peer Assessment 2

Jennifer Dimaano

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

Synopsis

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.

Data Processing

Libraries

```
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(knitr)
library(markdown)
library(rmarkdown)
library(lattice)
knitr::opts_chunk$set(error = TRUE)
```

Loading Data

```
if(!exists("storm.data")) {
    storm.data <- read.csv(bzfile("repdata_data_StormData.csv.bz2"),header = TRUE)
  }
```

Data Structure

```
dim(storm.data)
```

```
## [1] 902297     37
```

```
str(storm.data)
```

```
## 'data.frame':    902297 obs. of  37 variables:
##  $ STATE__   : num  1 1 1 1 1 1 1 1 1 1 ...
##  $ BGN_DATE  : Factor w/ 16335 levels "1/1/1966 0:00:00",..: 6523 6523 4242 11116 2224 2224 2260 383
##  $ BGN_TIME  : Factor w/ 3608 levels "00:00:00 AM",..: 272 287 2705 1683 2584 3186 242 1683 3186 318
##  $ TIME_ZONE : Factor w/ 22 levels "ADT","AKS","AST",..: 7 7 7 7 7 7 7 7 7 7 ...
##  $ COUNTY    : num  97 3 57 89 43 77 9 123 125 57 ...
##  $ COUNTYNAME: Factor w/ 29601 levels "","5NM E OF MACKINAC BRIDGE TO PRESQUE ISLE LT MI",..: 13513 1
##  $ STATE     : Factor w/ 72 levels "AK","AL","AM",..: 2 2 2 2 2 2 2 2 2 2 ...
##  $ EVTYPE    : Factor w/ 985 levels "   HIGH SURF ADVISORY",..: 834 834 834 834 834 834 834 834 834 8
##  $ BGN_RANGE : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ BGN_AZI   : Factor w/ 35 levels "","  N"," NW",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ BGN_LOCATI: Factor w/ 54429 levels "","  Christiansburg",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ END_DATE  : Factor w/ 6663 levels "","1/1/1993 0:00:00",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ END_TIME  : Factor w/ 3647 levels "","  0900CST",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ COUNTY_END: num  0 0 0 0 0 0 0 0 0 0 ...
##  $ COUNTYENDN: logi  NA NA NA NA NA NA ...
##  $ END_RANGE : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ END_AZI   : Factor w/ 24 levels "","E","ENE","ESE",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ END_LOCATI: Factor w/ 34506 levels "","  CANTON","  TULIA",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ LENGTH    : num  14 2 0.1 0 0 1.5 1.5 0 3.3 2.3 ...
##  $ WIDTH     : num  100 150 123 100 150 177 33 33 100 100 ...
##  $ F         : int  3 2 2 2 2 2 2 1 3 3 ...
##  $ MAG       : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ FATALITIES: num  0 0 0 0 0 0 0 0 1 0 ...
##  $ INJURIES  : num  15 0 2 2 2 6 1 0 14 0 ...
##  $ PROPDMG   : num  25 2.5 25 2.5 2.5 2.5 2.5 2.5 25 25 ...
##  $ PROPDMGEXP: Factor w/ 19 levels "","-","?","+",..: 17 17 17 17 17 17 17 17 17 17 ...
##  $ CROPDMG   : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ CROPDMGEXP: Factor w/ 9 levels "","?","0","2",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ WFO       : Factor w/ 542 levels "","  CI","%SD",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ STATEOFFIC: Factor w/ 250 levels "","ALABAMA, Central",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ ZONENAMES : Factor w/ 25112 levels "","
##  $ LATITUDE  : num  3040 3042 3340 3458 3412 ...
##  $ LONGITUDE : num  8812 8755 8742 8626 8642 ...
##  $ LATITUDE_E: num  3051 0 0 0 0 ...
##  $ LONGITUDE_: num  8806 0 0 0 0 ...
##  $ REMARKS   : Factor w/ 436781 levels "","\t","\t\t",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ REFNUM    : num  1 2 3 4 5 6 7 8 9 10 ...
```

Variables that will be used:

EVTYPE: Event Type (Tornados, Flood, ...)

FATALITIES: Number of Fatalities

INJURIES: Number of Injuries

PROGDMG: Property Damage

PROPDMGEXP: Units for Property Damage (magnitudes - K,B,M)

CROPDMG: Crop Damage

CROPDMGEXP: Units for Crop Damage (magnitudes - K,BM,B)

```
var <- c("EVTYPE", "FATALITIES", "INJURIES", "PROPDMG", "PROPDMGEXP", "CROPDMG", "CROPDMGEXP")
storm.data <- storm.data[var]
dim(storm.data)
```

```
## [1] 902297      7
```

checking data

```
head(storm.data)
```

```
##    EVTYPE FATALITIES INJURIES PROPDMG PROPDMGEXP CROPDMG CROPDMGEXP
## 1 TORNADO          0       15    25.0          K       0
## 2 TORNADO          0        0     2.5          K       0
## 3 TORNADO          0        2    25.0          K       0
## 4 TORNADO          0        2     2.5          K       0
## 5 TORNADO          0        2     2.5          K       0
## 6 TORNADO          0        6     2.5          K       0
```

```
tail(storm.data)
```

```
##                EVTYPE FATALITIES INJURIES PROPDMG PROPDMGEXP CROPDMG CROPDMGEXP
## 902292 WINTER WEATHER          0        0       0          K       0          K
## 902293      HIGH WIND          0        0       0          K       0          K
## 902294      HIGH WIND          0        0       0          K       0          K
## 902295      HIGH WIND          0        0       0          K       0          K
## 902296       BLIZZARD          0        0       0          K       0          K
## 902297     HEAVY SNOW          0        0       0          K       0          K
```

```
# NAs in variables
sum(is.na(storm.data$EVTYPE))
```

```
## [1] 0
```

```
sum(is.na(storm.data$FATALITIES))
```

```
## [1] 0
```

```
sum(is.na(storm.data$INJURIES))
```

```
## [1] 0
```

```
sum(is.na(storm.data$PROGDMG))
```

```
## [1] 0
```

```r
sum(is.na(storm.data$PROPDMGEXP))
```

```
## [1] 0
```

```r
sum(is.na(storm.data$CROPDMG))
```

```
## [1] 0
```

```r
sum(is.na(storm.data$CROPDMGEXP))
```

```
## [1] 0
```

```r
# checking dollar amounts
sort(table(storm.data$PROPDMGEXP), decreasing = TRUE)[1:10]
```

```
##
##             K      M      0      B      5      1      2      ?      m
## 465934 424665  11330    216     40     28     25     13      8      7
```

```r
sort(table(storm.data$CROPDMGEXP), decreasing = TRUE)[1:10]
```

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

```r
# checking top 10 events
sort(table(storm.data$EVTYPE), decreasing = TRUE)[1:10]
```

```
##
##               HAIL          TSTM WIND  THUNDERSTORM WIND            TORNADO
##             288661             219940              82563              60652
##        FLASH FLOOD              FLOOD THUNDERSTORM WINDS          HIGH WIND
##              54277              25326              20843              20212
##          LIGHTNING         HEAVY SNOW
##              15754              15708
```

Transforming variables

Group rest of values into the top 10 events by using keyword association. Events not associated with the top 10 keywords will be recorede as "OTHER." New variable EVENT is will hold the transformed records

```r
# create a new variable EVENT to transform variable EVTYPE in groups
storm.data$EVENT <- "OTHER"
# group by keyword in EVTYPE
storm.data$EVENT[grep("HAIL", storm.data$EVTYPE, ignore.case = TRUE)] <- "HAIL"
storm.data$EVENT[grep("HEAT", storm.data$EVTYPE, ignore.case = TRUE)] <- "HEAT"
storm.data$EVENT[grep("FLOOD", storm.data$EVTYPE, ignore.case = TRUE)] <- "FLOOD"
storm.data$EVENT[grep("WIND", storm.data$EVTYPE, ignore.case = TRUE)] <- "WIND"
storm.data$EVENT[grep("STORM", storm.data$EVTYPE, ignore.case = TRUE)] <- "STORM"
```

```
storm.data$EVENT[grep("SNOW", storm.data$EVTYPE, ignore.case = TRUE)] <- "SNOW"
storm.data$EVENT[grep("TORNADO", storm.data$EVTYPE, ignore.case = TRUE)] <- "TORNADO"
storm.data$EVENT[grep("WINTER", storm.data$EVTYPE, ignore.case = TRUE)] <- "WINTER"
storm.data$EVENT[grep("RAIN", storm.data$EVTYPE, ignore.case = TRUE)] <- "RAIN"
# listing the transformed event types
sort(table(storm.data$EVENT), decreasing = TRUE)
```

```
##
##    HAIL    WIND   STORM   FLOOD TORNADO   OTHER  WINTER    SNOW    RAIN    HEAT
##  289270  255362  113156   82686   60700   48970   19604   17660   12241    2648
```

Units need to be transformed. Below are the dollar conventions from NOAA: * K or k: thousand dollars (10^3) * M or m: million dollars (10^6) * B or b: billion dollars (10^9) * the rest would be consider as dollars

New variable(s) is product of value of damage and dollar unit.

```
storm.data$PROPDMGEXP <- as.character(storm.data$PROPDMGEXP)
storm.data$PROPDMGEXP[is.na(storm.data$PROPDMGEXP)] <- 0 # NA's considered as dollars
storm.data$PROPDMGEXP[!grepl("K|M|B", storm.data$PROPDMGEXP, ignore.case = TRUE)] <- 0 # everything exe
storm.data$PROPDMGEXP[grep("K", storm.data$PROPDMGEXP, ignore.case = TRUE)] <- "3"
storm.data$PROPDMGEXP[grep("M", storm.data$PROPDMGEXP, ignore.case = TRUE)] <- "6"
storm.data$PROPDMGEXP[grep("B", storm.data$PROPDMGEXP, ignore.case = TRUE)] <- "9"
storm.data$PROPDMGEXP <- as.numeric(as.character(storm.data$PROPDMGEXP))
storm.data$property.damage <- storm.data$PROPDMG * 10^storm.data$PROPDMGEXP

storm.data$CROPDMGEXP <- as.character(storm.data$CROPDMGEXP)
storm.data$CROPDMGEXP[is.na(storm.data$CROPDMGEXP)] <- 0 # NA's considered as dollars
storm.data$CROPDMGEXP[!grepl("K|M|B", storm.data$CROPDMGEXP, ignore.case = TRUE)] <- 0 # everything exe
storm.data$CROPDMGEXP[grep("K", storm.data$CROPDMGEXP, ignore.case = TRUE)] <- "3"
storm.data$CROPDMGEXP[grep("M", storm.data$CROPDMGEXP, ignore.case = TRUE)] <- "6"
storm.data$CROPDMGEXP[grep("B", storm.data$CROPDMGEXP, ignore.case = TRUE)] <- "9"
storm.data$CROPDMGEXP <- as.numeric(as.character(storm.data$CROPDMGEXP))
storm.data$crop.damage <- storm.data$CROPDMG * 10^storm.data$CROPDMGEXP
```

Analysis

Economic impact analysis

```
storm.data.damage <- storm.data %>%
  select(EVENT, property.damage,crop.damage) %>%
  group_by(EVENT) %>%
  summarise(total.dmg = sum(property.damage,crop.damage)) %>%
  arrange(-total.dmg)
head(storm.data.damage)
```

```
## # A tibble: 6 x 2
##   EVENT       total.dmg
##   <chr>           <dbl>
## 1 FLOOD   179769100029.
## 2 OTHER   120835593207.
## 3 STORM    72678890281.
```

```
## 4 TORNADO  59010559549.
## 5 HAIL     18779880521.
## 6 WIND     12250885768.
```

Public health impact analysis

```
# Fatalities
storm.data.fatalities <- storm.data %>%
  select(EVENT, FATALITIES) %>%
  group_by(EVENT) %>%
  summarise(total.fatalities = sum(FATALITIES)) %>%
  arrange(-total.fatalities)
head(storm.data.fatalities, 10)
```

```
## # A tibble: 10 x 2
##    EVENT   total.fatalities
##    <chr>              <dbl>
##  1 TORNADO             5661
##  2 HEAT                3138
##  3 OTHER               2626
##  4 FLOOD               1524
##  5 WIND                1209
##  6 STORM                416
##  7 WINTER               278
##  8 SNOW                 164
##  9 RAIN                 114
## 10 HAIL                  15
```

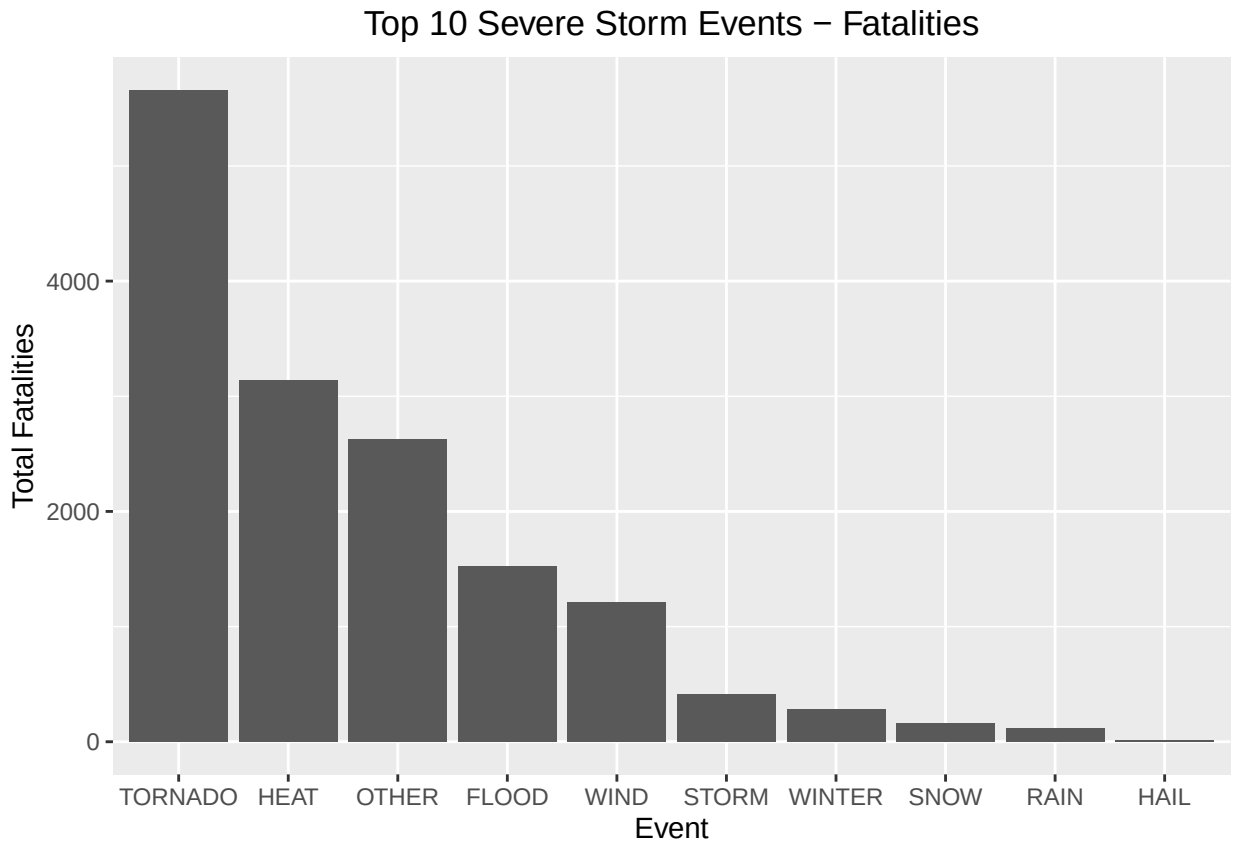
```
# Injuries
storm.data.injuries <- storm.data %>% select(EVENT, INJURIES) %>% group_by(EVENT) %>% summarise(total.i
head(storm.data.injuries, 10)
```

```
## # A tibble: 10 x 2
##    EVENT   total.injuries
##    <chr>            <dbl>
##  1 TORNADO          91407
##  2 OTHER            12224
##  3 HEAT              9224
##  4 WIND              9001
##  5 FLOOD             8602
##  6 STORM             5339
##  7 WINTER            1891
##  8 HAIL              1371
##  9 SNOW              1164
## 10 RAIN               305
```

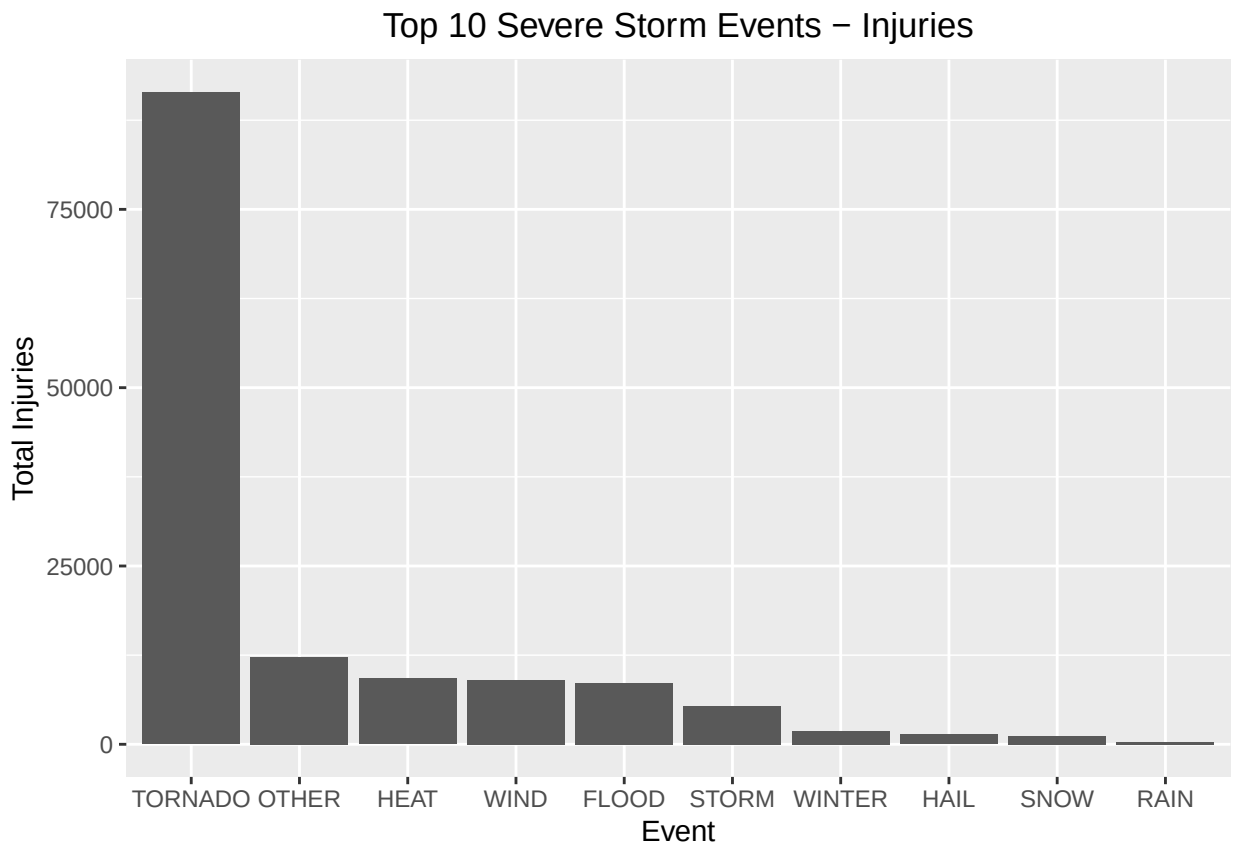
Results

```
Fatalities <- ggplot(storm.data.fatalities, aes(x=reorder(EVENT, -total.fatalities), y=total.fatalities
  geom_bar(stat = "identity") +
  xlab("Event") +
```

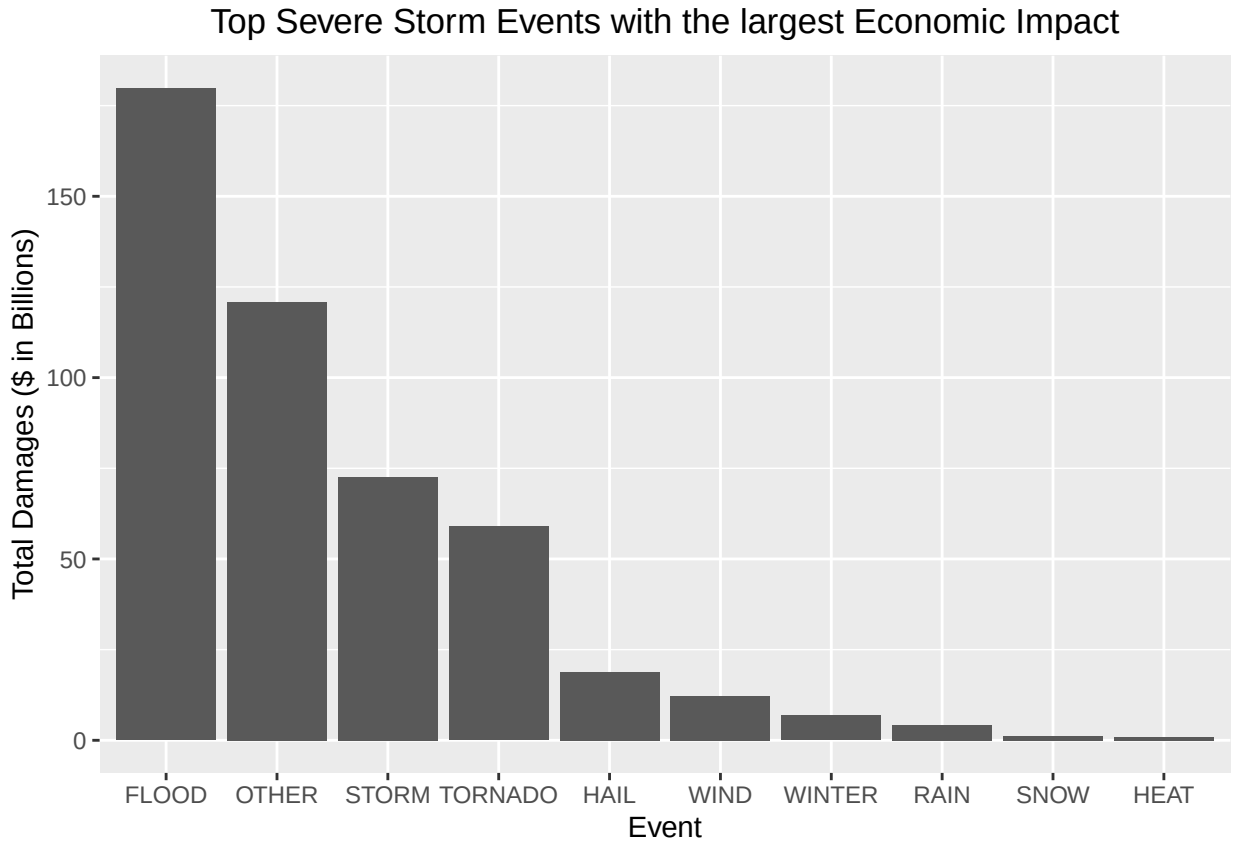
```
  ylab("Total Fatalities") +
  ggtitle("Top 10 Severe Storm Events - Fatalities") +
  theme(plot.title = element_text(hjust = 0.5))

Injuries <- ggplot(storm.data.injuries, aes(x=reorder(EVENT, -total.injuries), y=total.injuries)) +
  geom_bar(stat = "identity") +
  xlab("Event") +
  ylab("Total Injuries") +
  ggtitle("Top 10 Severe Storm Events - Injuries") +
  theme(plot.title = element_text(hjust = 0.5))

Damages <- ggplot(storm.data.damage, aes(x=reorder(EVENT, -total.dmg), y=(total.dmg/1000000000))) +
  geom_bar(stat = "identity") +
  xlab("Event") +
  ylab("Total Damages ($ in Billions)") +
  ggtitle("Top Severe Storm Events with the largest Economic Impact") +
  theme(plot.title = element_text(hjust = 0.5))

print(Fatalities)
```




```
print(Injuries)
```




```
print(Damages)
```



Tornados make up the largest public health impact resulting in 91,407 annual injuries and 5,661 annual fatalities. Heat and other severe storm events follow after Tornados. Hail makes up the least amount of fatalities at 15, while rain produces only 305 injuries over a year.

In terms of economic impact, flood damaage resulted in over $179B. While Tornados made up a fraction of that damage at $5B.