Project 2_storm data

The impacts of natural disasters on U.S. population health and econoic consequences from 1950s to 1990s

weimf1990

12/08/2021

Synopsis

This project involves exploring the U.S. National Oceanic and Atmospheric Administration's (NOAA) storm database. I intend to analyze what types of natural disasters have greater threats on population health and economic consequences from 1950s to 1990s. In the main, I find that tornado most harmful to population health in terms of injuries and fatalities cases. On the other hand, flood contributes to severe economic consequences the most.

Data Processing

Loading the data

Data source: Storm data (https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2)
Documentation of the database available: National Weather Service Instruction (https://d396qusza40orc.cloudfront.net/repdata%2Fpeer2_doc%2Fpd01016005curr.pdf)

```
#setting the directory
setwd("C:/Users/user/Desktop/Data Science/project/Course 5")

#loading the data
stormdata <- read.csv("./repdata_data_StormData.csv.bz2")

#check the size of the data
dim(stormdata)
```

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

There are more than 90k observations and 37 variables in the raw daa. To make the analysis more efficient, I need to process the data and keep only information required for the work.

For the purpose of this analysis, only observations with `injuries or damage value >= 0` will be included. Thus, I drop the missing values without replacement. I do not intend to choose any method to impute the missing vlaues in the dataet to maintain the originality of the data.

Variables that are used for this analysis are:
- CROPDMG : Crop damage measured
- CROPDMGEXP : Crop damage exponent (K, M, B, etc.)
- EVTYPE : Event type (Abnormal Warmth, Astronomical Low Tide, Tornado…)

- FATALITIES : Number of fatalities from the event
- INJURIES : Nunber of injuries from the event
- PROPDMG : Property damage measured
- PROPDMGEXP : Property damage exponent (K, M, B, etc.)
- BGN_DATE: Begin date of the event
- END_DATE: End date of the event
- STATE: State where the event occurred

```
final_stormdata <- subset(stormdata, (FATALITIES > 0 | INJURIES > 0 | PROPDMG > 0 | CROPDMG > 0
),
                                   select = c("EVTYPE",
                                              "FATALITIES",
                                              "INJURIES",
                                              "PROPDMG",
                                              "PROPDMGEXP",
                                              "CROPDMG",
                                              "CROPDMGEXP",
                                              "BGN_DATE",
                                              "END_DATE",
                                              "STATE")
                          )
dim(final_stormdata)
```

```
## [1] 254633     10
```

```
dim(final_stormdata)
```

```
## [1] 254633     10
```

```
sum(is.na(final_stormdata))
```

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

The alphabetical characters are used to signify magnitude, such as a “K” for thousands, “M” for millions, and “B” for billions.

```
#view the variables with alphabetical characters
unique(final_stormdata$CROPDMGEXP)
unique(final_stormdata$PROPDMGEXP)
unique(final_stormdata$EVTYPE)
```

However, the variables “CROPDMGEXP” and “PROPDMGEXP” need to be reorganized to measure the damages in numeric values.

Cleaning the variables with exponential values and creating the numeric damage costs

```
# create a function to get the multiplier factor
Exponentvalues <- function(exp) {

  #use "toupper" method to change the case to upper and generate the variable "exp"
  exp <- toupper(exp);

  # "+", "-", "?", " ", "0" -> 1
  if(exp == "+") return (10^0);
  if(exp == "-") return (10^0);
  if(exp == "?") return (10^0);
  if(exp == " ") return (10^0);
  if(exp == "0") return (10^0);

  # 1-9 -> 10^(number)
  if(exp == "1") return (10^1);
  if(exp == "2") return (10^2);
  if(exp == "3") return (10^3);
  if(exp == "4") return (10^4);
  if(exp == "5") return (10^5);
  if(exp == "6") return (10^6);
  if(exp == "7") return (10^7);
  if(exp == "8") return (10^8);
  if(exp == "9") return (10^9);

  # H,h -> 100
  if(exp == "H") return (10^2);
  if(exp == "h") return (10^2);

  # K,k -> 1,000
  if(exp == "K") return (10^3);
  if(exp == "k") return (10^3);

  # M,m -> 1,000,000
  if(exp == "M") return (10^6);
  if(exp == "m") return (10^6);

  # B,b -> 1,000,000,000
  if(exp == "B") return (10^9);
  if(exp == "b") return (10^9);

  #something else
  return (NA)
}

# generate numberic damage values (in billions), and the valuables are named "cropdamagecost" an
d "propdamagecost", respectively
final_stormdata$cropdamagecost <- with(final_stormdata, as.numeric(CROPDMG) * sapply(CROPDMGEXP,
Exponentvalues)/(10^9))

final_stormdata$propdamagecost <- with(final_stormdata, as.numeric(PROPDMG) * sapply(PROPDMGEXP,
Exponentvalues)/(10^9))
```

Organizing event type information

So does the Event Type data. To make the values consistent, I reorganize all Event Type values to uppercase. These categories should fit into the listed event types in section 7 of the document (https://d396qusza40orc.cloudfront.net/repdata%2Fpeer2_doc%2Fpd01016005curr.pdf). I leave other uncategorized items for now because they are less likely to be the most severe event I intend to analyze in this report.

```
#removing all white space and making all characters upper case
final_stormdata$EVTYPE <- toupper(trimws(final_stormdata$EVTYPE))

# AVALANCHE
final_stormdata$EVTYPE <- gsub('.*AVALANCE.*', 'AVALANCHE', final_stormdata$EVTYPE)

# BLIZZARD
final_stormdata$EVTYPE <- gsub('.*BLIZZARD.*', 'BLIZZARD', final_stormdata$EVTYPE)

# COLD
final_stormdata$EVTYPE <- gsub('.*COLD.*', 'COLD', final_stormdata$EVTYPE)
final_stormdata$EVTYPE <- gsub('.*COOL.*', 'COLD', final_stormdata$EVTYPE)
final_stormdata$EVTYPE <- gsub('.*LOW TEMPERATURE RECORD.*', 'COLD', final_stormdata$EVTYPE)
final_stormdata$EVTYPE <- gsub('.*LO.*TEMP.*', 'COLD', final_stormdata$EVTYPE)

# FREEZE (FROST included)
final_stormdata$EVTYPE <- gsub('.*FREEZE.*', 'FREEZE', final_stormdata$EVTYPE)
final_stormdata$EVTYPE <- gsub('.*FROST.*', 'FREEZE', final_stormdata$EVTYPE)

# ICE
final_stormdata$EVTYPE <- gsub('.*ICE.*', 'ICE', final_stormdata$EVTYPE)
final_stormdata$EVTYPE <- gsub('.*ICY.*', 'ICE', final_stormdata$EVTYPE)

# DROUGHT
final_stormdata$EVTYPE <- gsub('.*DROUGHT.*', 'DROUGHT', final_stormdata$EVTYPE)
final_stormdata$EVTYPE <- gsub('.*DRY.*', 'DROUGHT', final_stormdata$EVTYPE)

# DUST
final_stormdata$EVTYPE <- gsub('.*DUST.*', 'DUST', final_stormdata$EVTYPE)

# FLOOD
final_stormdata$EVTYPE <- gsub('.*FLOOD.*', 'FLOOD', final_stormdata$EVTYPE)
final_stormdata$EVTYPE <- gsub('.*FLOODING.*', 'FLOOD', final_stormdata$EVTYPE)

# FOG
final_stormdata$EVTYPE <- gsub('.*FOG.*', 'FOG', final_stormdata$EVTYPE)

# HAIL
final_stormdata$EVTYPE <- gsub('.*HAIL.*', 'HAIL', final_stormdata$EVTYPE)

# HEAT
final_stormdata$EVTYPE <- gsub('.*HEAT.*', 'HEAT', final_stormdata$EVTYPE)
final_stormdata$EVTYPE <- gsub('.*HIGH.*TEMP.*', 'HEAT', final_stormdata$EVTYPE)
final_stormdata$EVTYPE <- gsub('.*RECORD HIGH TEMPERATURES.*', 'HEAT', final_stormdata$EVTYPE)

#HURRICANE (TYPHOON included)
final_stormdata$EVTYPE <- gsub('.*HURRICANE.*', 'HURRICANE', final_stormdata$EVTYPE)
final_stormdata$EVTYPE <- gsub('.*TYPHOON.*', 'HURRICANE', final_stormdata$EVTYPE)

# LIGHTNING
final_stormdata$EVTYPE <- gsub('^LIGHTNING.*', 'LIGHTNING', final_stormdata$EVTYPE)
```

```
# RAIN
final_stormdata$EVTYPE <- gsub('.*RAIN.*', 'RAIN', final_stormdata$EVTYPE)

# RIP CURRENT
final_stormdata$EVTYPE <- gsub('.*RIP CURRENT.*', 'RIP CURRENT', final_stormdata$EVTYPE)

# STORM
final_stormdata$EVTYPE <- gsub('.*STORM.*', 'STORM', final_stormdata$EVTYPE)

# SURF
final_stormdata$EVTYPE <- gsub('.*SURF.*', 'SURF', final_stormdata$EVTYPE)

# TORNADO
final_stormdata$EVTYPE <- gsub('.*TORNADO.*', 'TORNADO', final_stormdata$EVTYPE)

# VOLCANIC
final_stormdata$EVTYPE <- gsub('.*VOLCANIC.*', 'VOLCANIC', final_stormdata$EVTYPE)

#WATERSPOUT
final_stormdata$EVTYPE <- gsub('.*WATERSPOUT.*', 'WATERSPOUT', final_stormdata$EVTYPE)

# WET
final_stormdata$EVTYPE <- gsub('.*WET.*', 'WET', final_stormdata$EVTYPE)

# WIND
final_stormdata$EVTYPE <- gsub('.*WIND.*', 'WIND', final_stormdata$EVTYPE)

# WINTER
final_stormdata$EVTYPE <- gsub('.*WINTER.*', 'WINTER', final_stormdata$EVTYPE)
final_stormdata$EVTYPE <- gsub('.*SNOW.*', 'WINTER', final_stormdata$EVTYPE)

#WILDFIRE
final_stormdata$EVTYPE <- gsub('.*FIRE.*', 'WILDFIRE', final_stormdata$EVTYPE)
```

Results

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

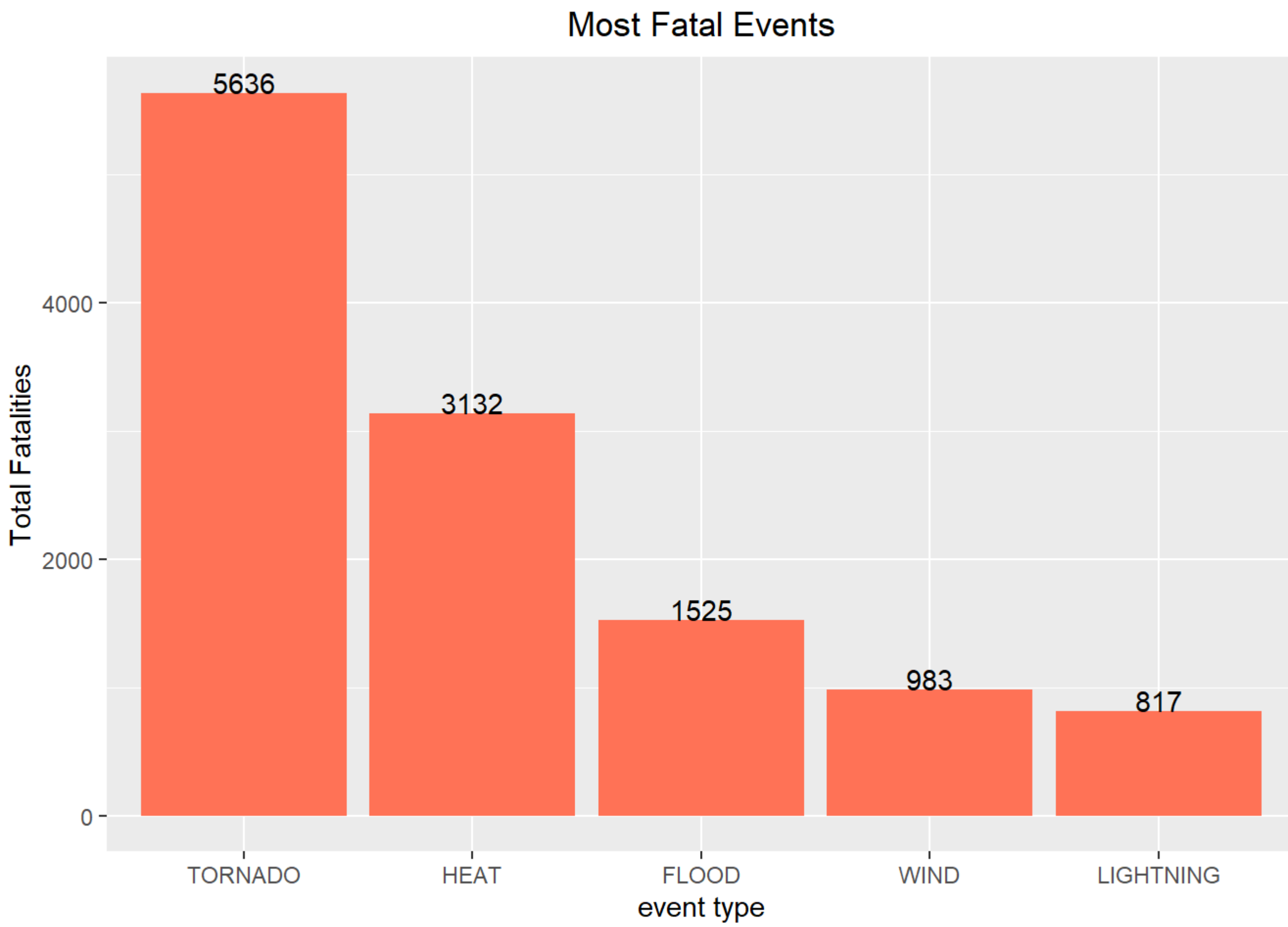
```
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 4.0.5
```

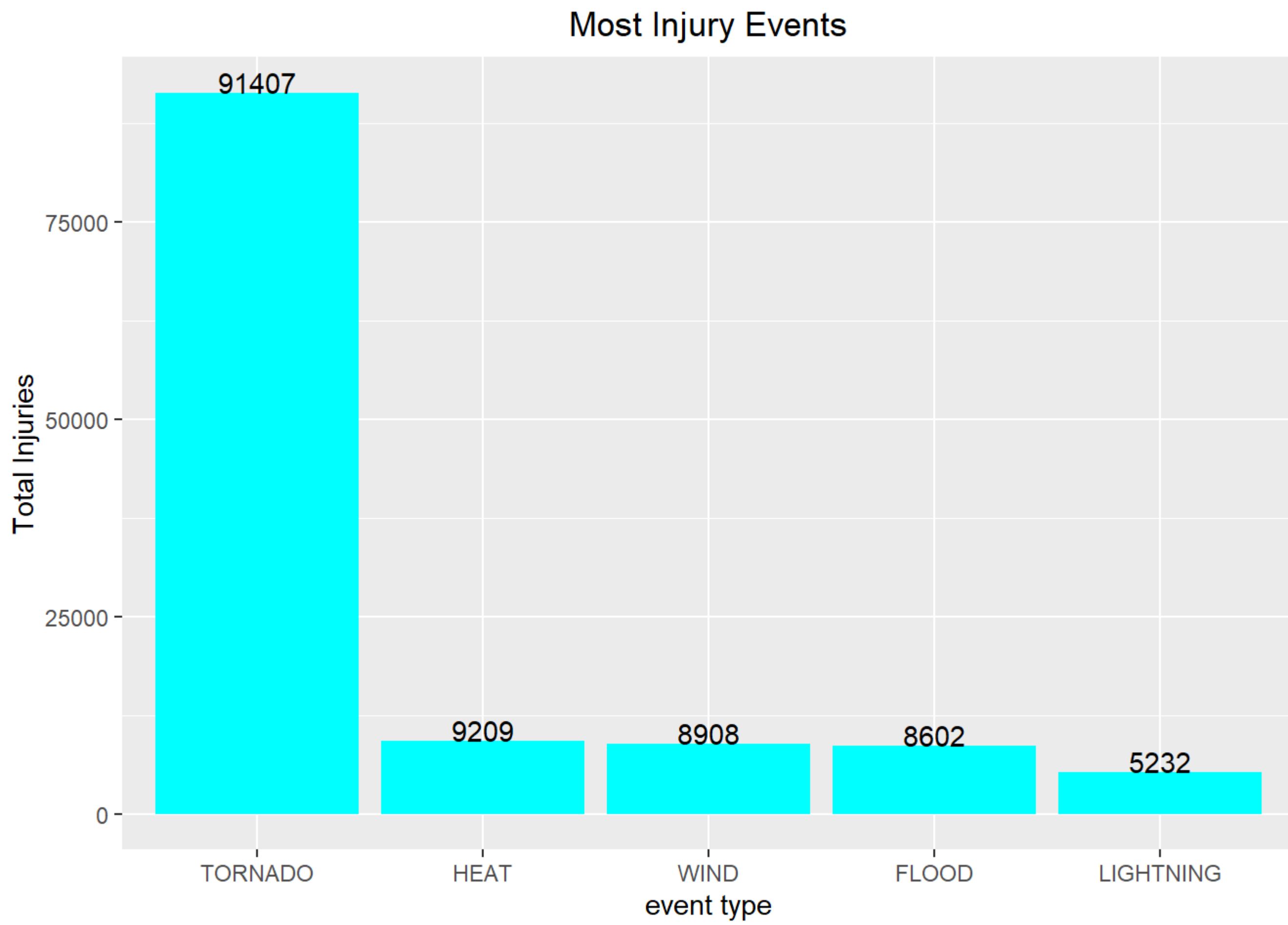
```
Summary_data <- final_stormdata %>%
    group_by(EVTYPE) %>%
    summarize(SUMFATALITIES = sum(FATALITIES),
              SUMINJURIES = sum(INJURIES),
              SUMHARMFUL = sum(FATALITIES, INJURIES),
              SUMPROPDMG = round(sum(propdamagecost, na.rm = TRUE), digits = 2),
              SUMCROPDMG = round(sum(cropdamagecost, na.rm = TRUE), digits = 2),
              TOTALDMG = round(sum(propdamagecost, cropdamagecost, na.rm = TRUE), digits = 2)
              )

#Plot 1: fatility
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.0.5
```

```
SumFatalities <- arrange(Summary_data, desc(SUMFATALITIES))
FatalityData <- head(SumFatalities, n=5)
FatalityData$EVTYPE <- with(FatalityData, reorder(EVTYPE, -SUMFATALITIES))

g1 <- ggplot(FatalityData, aes(EVTYPE, SUMFATALITIES)) +
    geom_bar(stat ="identity", fill = "coral1") +
    geom_text(aes(label=SUMFATALITIES), vjust=0) +
    labs(x = "event type") +
    labs(y = "Total Fatalities") +
    labs(title = "Most Fatal Events")+
    theme(plot.title = element_text(hjust = 0.5))

g1
```



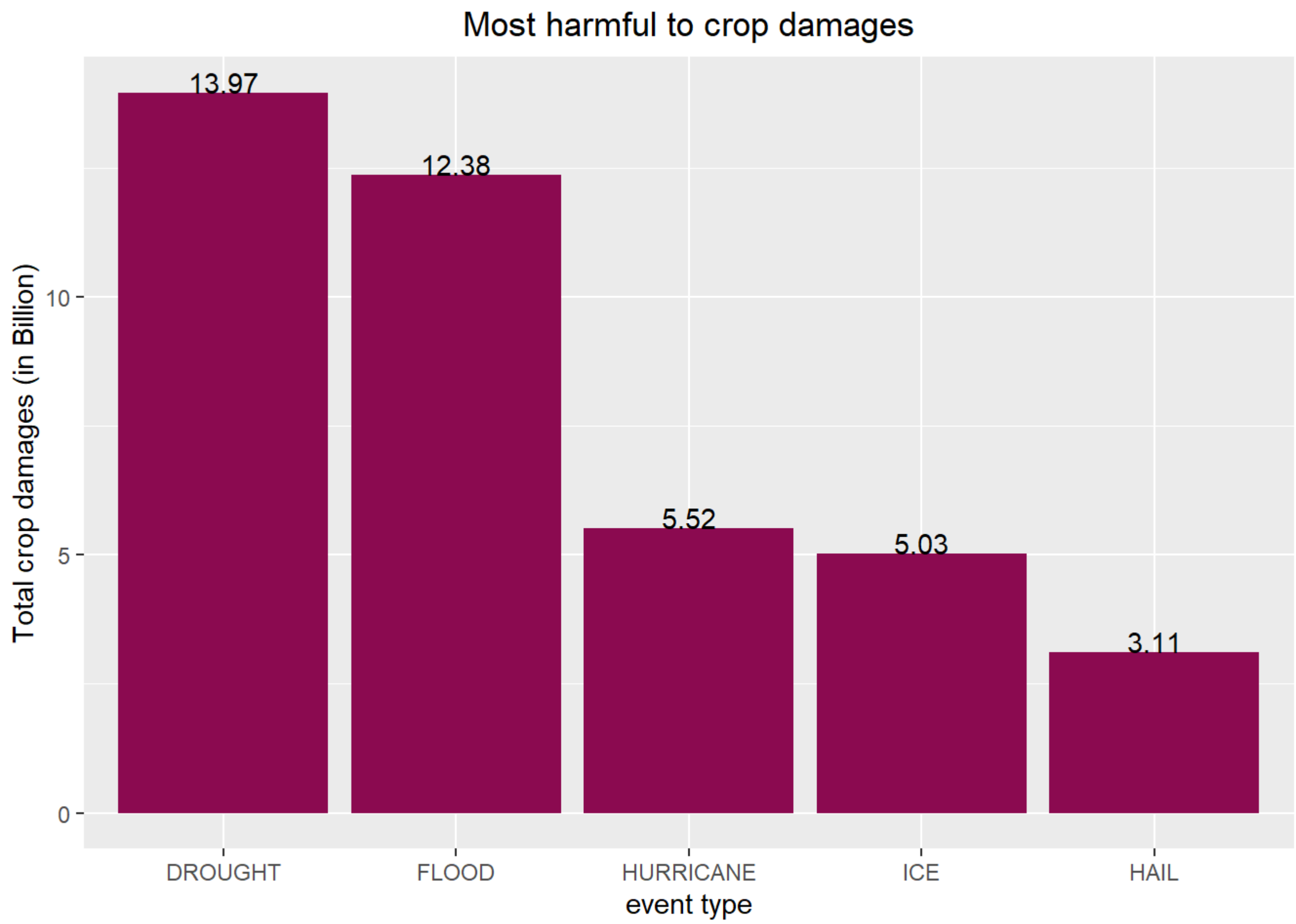

```
#Plot 2: injury
SumInjuries <- arrange(Summary_data, desc(SUMINJURIES))
InjuryData <- head(SumInjuries, n=5)
InjuryData$EVTYPE <- with(InjuryData, reorder(EVTYPE, -SUMINJURIES))

g2 <- ggplot(InjuryData, aes(EVTYPE, SUMINJURIES)) +
    geom_bar(stat ="identity", fill = "cyan1" ) +
    geom_text(aes(label=SUMINJURIES), vjust=0) +
    labs(x = "event type") +
    labs(y = "Total Injuries") +
    labs(title = "Most Injury Events")+
    theme(plot.title = element_text(hjust = 0.5))

g2
```



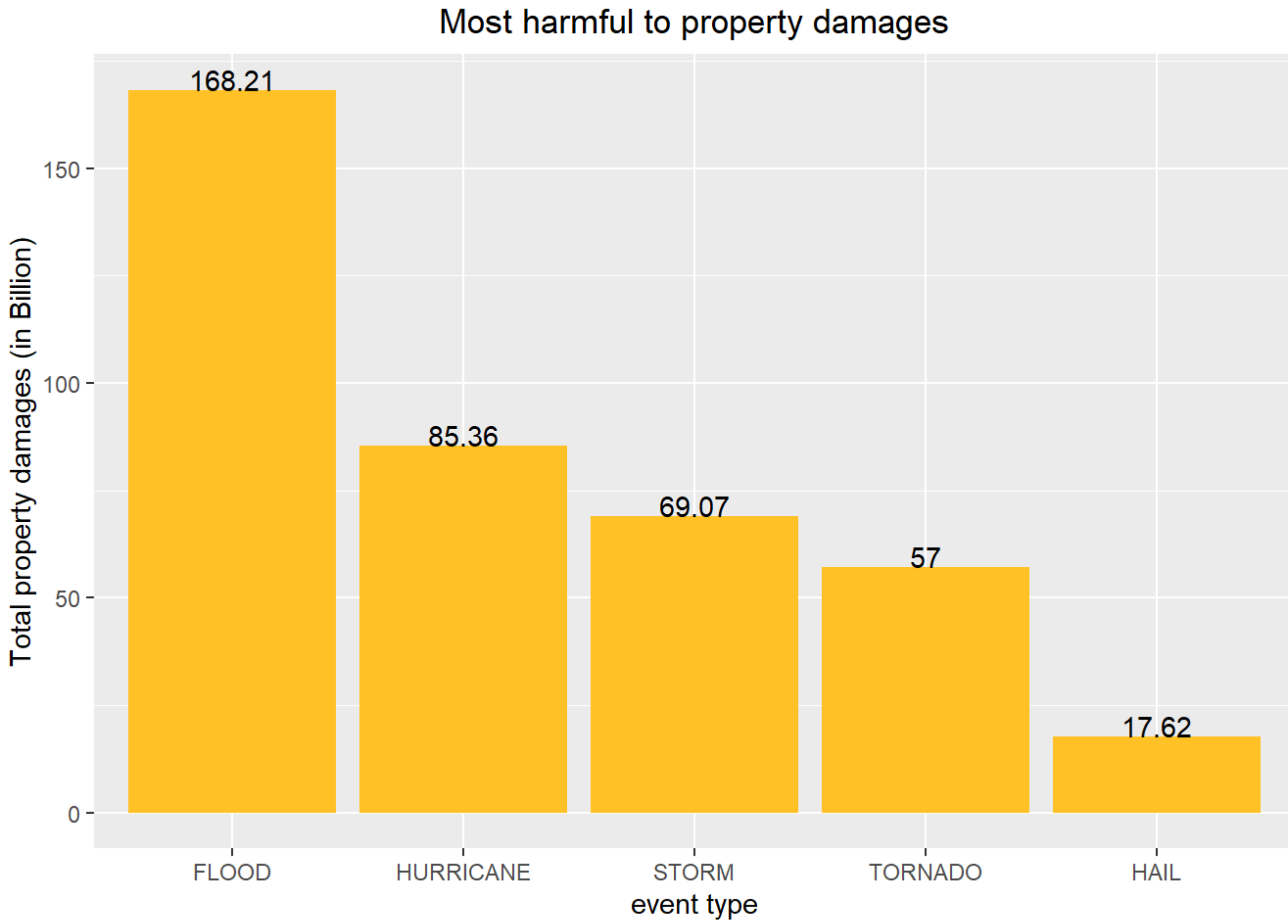
Over the years, tornado, heat, flood, wind, and lighting are the top five harmful events with respect to population health across the U.S. Specifically, tornado causes the most number of injuries and fatalities.

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

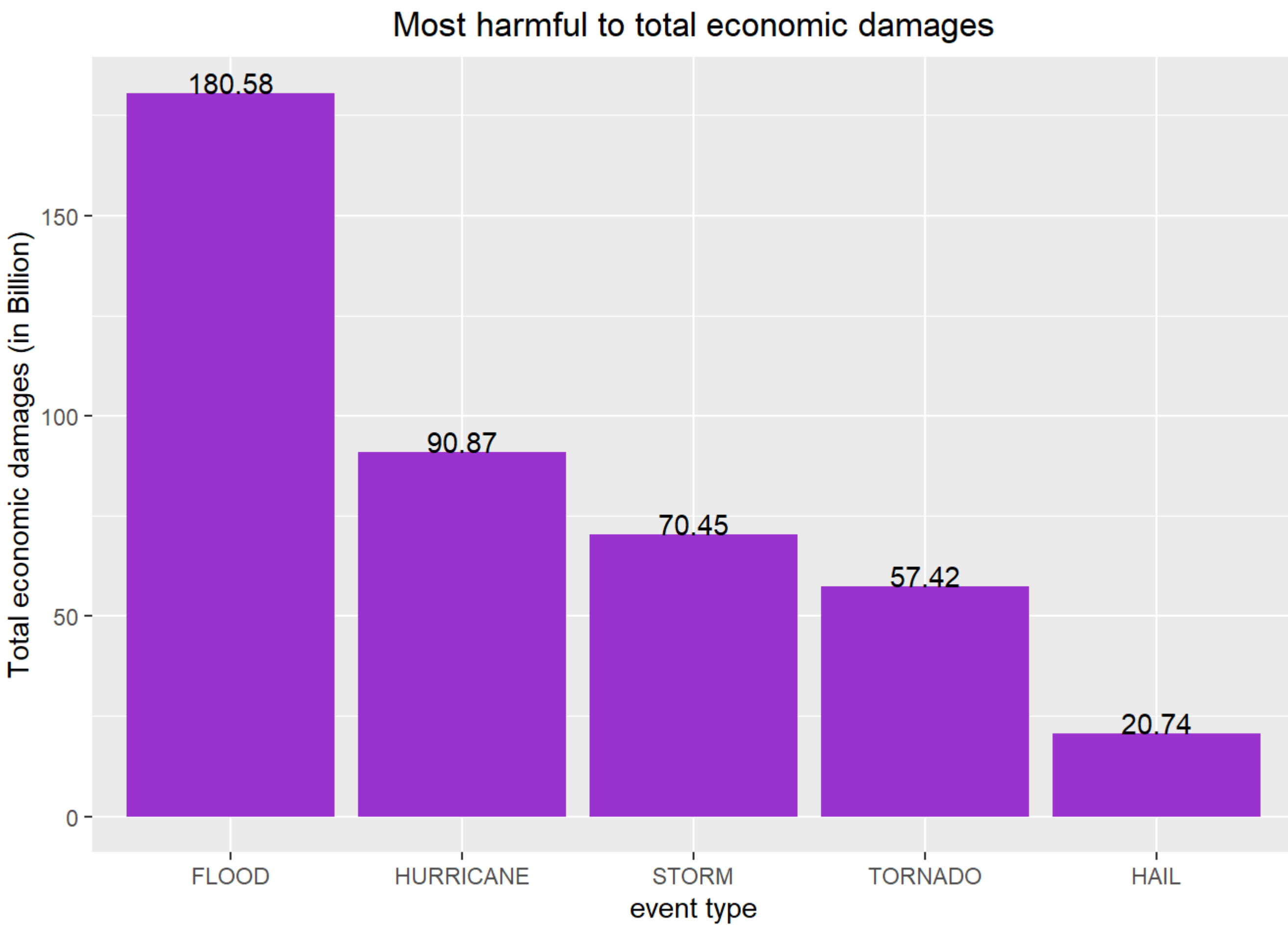
```
#plot 3: crop damage
Sumcroplost <- arrange(Summary_data, desc(SUMCROPDMG))
CropData <- head(Sumcroplost, n=5)
CropData$EVTYPE <- with(CropData, reorder(EVTYPE, -SUMCROPDMG))

g3 <- ggplot(CropData, aes(EVTYPE, SUMCROPDMG)) +
    geom_bar(stat ="identity", fill = "deeppink4") +
    geom_text(aes(label=SUMCROPDMG), vjust=0) +
    labs(x = "event type") +
    labs(y = "Total crop damages (in Billion)") +
    labs(title = "Most harmful to crop damages")+
    theme(plot.title = element_text(hjust = 0.5))

g3
```




```
#Plot 4: property damage
Sumpropertylost <- arrange(Summary_data, desc(SUMPROPDMG))
PropertyData <- head(Sumpropertylost, n=5)
PropertyData$EVTYPE <- with(PropertyData, reorder(EVTYPE, -SUMPROPDMG))

g4 <- ggplot(PropertyData, aes(EVTYPE, SUMPROPDMG)) +
    geom_bar(stat ="identity", fill = "goldenrod1") +
    geom_text(aes(label=SUMPROPDMG), vjust=0) +
    labs(x = "event type") +
    labs(y = "Total property damages (in Billion)") +
    labs(title = "Most harmful to property damages")+
    theme(plot.title = element_text(hjust = 0.5))

g4
```




```
#Plot 5: total damage
Sumtotallost <- arrange(Summary_data, desc(TOTALDMG))
TotaldamageData <- head(Sumtotallost, n=5)
TotaldamageData$EVTYPE <- with(TotaldamageData, reorder(EVTYPE, -TOTALDMG))

g5 <- ggplot(TotaldamageData, aes(EVTYPE, TOTALDMG)) +
    geom_bar(stat ="identity", fill = "darkorchid3") +
    geom_text(aes(label=TOTALDMG), vjust=0) +
    labs(x = "event type") +
    labs(y = "Total economic damages (in Billion)") +
    labs(title = "Most harmful to total economic damages")+
    theme(plot.title = element_text(hjust = 0.5))

g5
```



Conclusions

Regarding the impact of natural disasters on economic consequences, drought and flood are the top two harmful events to crop damages. However, water-related disasters including flood, hurricane and storm hurt a higher extent on the property loss. Overall, flood causes the primary economic loss to the nation.