Influence of severe weather events on public health and economics

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 Furthermore, this document was prepared on the frame of the Peer Assessment 2 Reproducible Research Coursera Course, on 2014/07/27 by MonicaPH.

The effects of weather on public health and economics are examined, especifically the following questions: 1. Across the United States, which types of events are most harmful with respect to population health? 2. Across the United States, which types of events have the greatest economic consequences?

The natural disaster that casuses the most Population Health damages are tornados, followed by heat and floods. On the other hand, the events that cause the most economic losses are floods followed by tornados.

Data

To answer the proposed questions, data from the U.S. National Oceanic and Atmospheric Administration's (NOAA) storm database was used, as downloaded from the Coursera platform (https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2). 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. The raw data come in the form of a comma-separated-value file compressed via the bzip2 algorithm.

Extra information about the data set can be found at:

- National Weather Service Storm Data Documentation (https://d396qusza40orc.cloudfront.net/repdata%2Fpeer2_doc%2Fpd01016005curr.pdf)
- National Climatic Data Center Storm Events FAQ (https://d396qusza40orc.cloudfront.net/repdata%2Fpeer2_doc%2FNCDC%20Storm%20Events-FAQ%20Page.pdf)

Data Processing

Firts it loads the required libraries. Then instructions are included to download, unzip and read the data:

```
#Libraries
library(plyr)
library(ggplot2)



Sys.setlocale("LC_TIME", "English")       # Set language to engish
```

```
## [1] "English_United States.1252"
```

```r
#Downloading
fileurl<-"http://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2"
if(!file.exists("StormData.csv.bz2")){
      download.file(fileurl, destfile="StormData.csv.bz2")
}

#Reading
if(file.exists("StormData.csv.bz2")){
      data <- bzfile("StormData.csv.bz2", "r")
      stormD <- read.csv(data)
      close(data)
}

# Summary of the Data
summary(stormD)
```

```
##     STATE__                  BGN_DATE             BGN_TIME     
##  Min.   : 1.0   5/25/2011 0:00:00:  1202   12:00:00 AM: 10163  
##  1st Qu.:19.0   4/27/2011 0:00:00:  1193   06:00:00 PM:  7350  
##  Median :30.0   6/9/2011 0:00:00 :  1030   04:00:00 PM:  7261  
##  Mean   :31.2   5/30/2004 0:00:00:  1016   05:00:00 PM:  6891  
##  3rd Qu.:45.0   4/4/2011 0:00:00 :  1009   12:00:00 PM:  6703  
##  Max.   :95.0   4/2/2006 0:00:00 :   981   03:00:00 PM:  6700  
##                 (Other)          :895866   (Other)    :857229  
##    TIME_ZONE          COUNTY           COUNTYNAME         STATE       
##  CST    :547493   Min.   :  0   JEFFERSON :  7840   TX     : 83728  
##  EST    :245558   1st Qu.: 31   WASHINGTON:  7603   KS     : 53440  
##  MST    : 68390   Median : 75   JACKSON   :  6660   OK     : 46802  
##  PST    : 28302   Mean   :101   FRANKLIN  :  6256   MO     : 35648  
##  AST    :  6360   3rd Qu.:131   LINCOLN   :  5937   IA     : 31069  
##  HST    :  2563   Max.   :873   MADISON   :  5632   NE     : 30271  
##  (Other):  3631                 (Other)   :862369   (Other):621339  
##                EVTYPE          BGN_RANGE        BGN_AZI      
##  HAIL             :288661   Min.   :   0          :547332  
##  TSTM WIND        :219940   1st Qu.:   0   N      : 86752  
##  THUNDERSTORM WIND: 82563   Median :   0   W      : 38446  
##  TORNADO          : 60652   Mean   :   1   S      : 37558  
##  FLASH FLOOD      : 54277   3rd Qu.:   1   E      : 33178  
##  FLOOD            : 25326   Max.   :3749   NW     : 24041  
##  (Other)          :170878                  (Other):134990  
##          BGN_LOCATI                  END_DATE             END_TIME     
##               :287743                    :243411              :238978  
##  COUNTYWIDE   : 19680   4/27/2011 0:00:00:  1214   06:00:00 PM:  9802  
##  Countywide   :   993   5/25/2011 0:00:00:  1196   05:00:00 PM:  8314  
##  SPRINGFIELD  :   843   6/9/2011 0:00:00 :  1021   04:00:00 PM:  8104  
##  SOUTH PORTION:   810   4/4/2011 0:00:00 :  1007   12:00:00 PM:  7483  
##  NORTH PORTION:   784   5/30/2004 0:00:00:   998   11:59:00 PM:  7184  
##  (Other)      :591444   (Other)          :653450   (Other)    :622432  
##    COUNTY_END COUNTYENDN       END_RANGE       END_AZI      
```

```
##  Min.   :0    Mode:logical   Min.   :  0          :724837  
##  1st Qu.:0    NA's:902297    1st Qu.:  0   N      : 28082  
##  Median :0                   Median :  0   S      : 22510  
##  Mean   :0                   Mean   :  1   W      : 20119  
##  3rd Qu.:0                   3rd Qu.:  0   E      : 20047  
##  Max.   :0                   Max.   :925   NE     : 14606  
##                                            (Other): 72096  
##            END_LOCATI         LENGTH              WIDTH            F         
##                 :499225   Min.   :   0.0   Min.   :   0   Min.   :0       
##  COUNTYWIDE     : 19731   1st Qu.:   0.0   1st Qu.:   0   1st Qu.:0       
##  SOUTH PORTION  :   833   Median :   0.0   Median :   0   Median :1       
##  NORTH PORTION  :   780   Mean   :   0.2   Mean   :   8   Mean   :1       
##  CENTRAL PORTION:   617   3rd Qu.:   0.0   3rd Qu.:   0   3rd Qu.:1       
##  SPRINGFIELD    :   575   Max.   :2315.0   Max.   :4400   Max.   :5       
##  (Other)        :380536                                   NA's   :843563  
##       MAG            FATALITIES     INJURIES            PROPDMG       
##  Min.   :    0   Min.   :  0   Min.   :   0.0   Min.   :   0  
##  1st Qu.:    0   1st Qu.:  0   1st Qu.:   0.0   1st Qu.:   0  
##  Median :   50   Median :  0   Median :   0.0   Median :   0  
##  Mean   :   47   Mean   :  0   Mean   :   0.2   Mean   :  12  
##  3rd Qu.:   75   3rd Qu.:  0   3rd Qu.:   0.0   3rd Qu.:   0  
##  Max.   :22000   Max.   :583   Max.   :1700.0   Max.   :5000  
##                                                               
##    PROPDMGEXP        CROPDMG          CROPDMGEXP          WFO        
##         :465934   Min.   :  0.0          :618413          :142069  
##  K      :424665   1st Qu.:  0.0   K      :281832   OUN    : 17393  
##  M      : 11330   Median :  0.0   M      :  1994   JAN    : 13889  
##  0      :   216   Mean   :  1.5   k      :    21   LWX    : 13174  
##  B      :    40   3rd Qu.:  0.0   0      :    19   PHI    : 12551  
##  5      :    28   Max.   :990.0   B      :     9   TSA    : 12483  
##  (Other):    84                   (Other):     9   (Other):690738  
##                                STATEOFFIC    
##                                     :248769  
##  TEXAS, North                       : 12193  
##  ARKANSAS, Central and North Central: 11738  
##  IOWA, Central                      : 11345  
##  KANSAS, Southwest                  : 11212  
##  GEORGIA, North and Central         : 11120  
##  (Other)                            :595920  

              ZONENAMES
##

                   :594029
##

                   :205988
##  GREATER RENO / CARSON CITY / M - GREATER RENO / CARSON CITY / M
```

```
                   :   639
##  GREATER LAKE TAHOE AREA - GREATER LAKE TAHOE AREA

                   :   592
##  JEFFERSON - JEFFERSON

                   :   303
##  MADISON - MADISON

                   :   302
##  (Other)

                   :100444
##     LATITUDE      LONGITUDE        LATITUDE_E     LONGITUDE_
##  Min.   :   0   Min.   :-14451   Min.   :   0   Min.   :-14455
##  1st Qu.:2802   1st Qu.:  7247   1st Qu.:   0   1st Qu.:     0
##  Median :3540   Median :  8707   Median :   0   Median :     0
##  Mean   :2875   Mean   :  6940   Mean   :1452   Mean   :  3509
##  3rd Qu.:4019   3rd Qu.:  9605   3rd Qu.:3549   3rd Qu.:  8735
##  Max.   :9706   Max.   : 17124   Max.   :9706   Max.   :106220
##  NA's   :47                      NA's   :40
##                                            REMARKS           REFNUM
##                                                :287433   Min.   :     1
##                                                : 24013   1st Qu.:225575
##  Trees down.\n                                 :  1110   Median :451149
##  Several trees were blown down.\n              :   568   Mean   :451149
##  Trees were downed.\n                          :   446   3rd Qu.:676723
##  Large trees and power lines were blown down.\n:   432   Max.   :902297
##  (Other)                                       :588295
```

The dataset has 37 variables describing storms and its consequences. The relevant variables are selected for further analysis, namely: 1. **BGN_DATE** - Begin date 2. **BGN_TIME** - Begin time 3. **EVTYPE** - Type of event 4. **FATALITIES** - Fatalities to humans caused by the event 5. **INJURIES** - Injuries to humans caused by the event 6. **PROPDMG** - Property damage 7. **PROPDMGEXP** - Order of the property damage 8. **CROPDMG** - Crop damage 9. **CROPDMGEXP** - Order of the crop damage

```
# Getting the relevant columns
stormDs <- stormD[,c("BGN_DATE","BGN_TIME","EVTYPE","FATALITIES","INJURIES","PROPDMG","PROPD
MGEXP","CROPDMG","CROPDMGEXP")]

# Formatting Date and Time together
stormDs[,10] <- paste(substr(stormDs[,1],1,10),stormDs[,2])
stormDs[,1] <- data.frame(strptime(stormDs[,10], format="%m/%d/%Y %H%M"))[,1]
stormDs <- stormDs[,c("BGN_DATE","EVTYPE","FATALITIES","INJURIES","PROPDMG","PROPDMGEXP","CR
OPDMG","CROPDMGEXP")]
```

The orders of magnitude are coded, therefore, they need to be translated and grouped into a single magnitude variable

```
# Check Unique order levels property damage
summary(stormDs[,6])
```

```
##              -      ?      +      0      1      2      3      4      5
## 465934      1      8      5    216     25     13      4      4     28
##      6      7      8      B      h      H      K      m      M
##      4      5      1     40      1      6 424665      7  11330
```

```
# Replace order of magnitude by numbers
levels(stormDs[,6])<-c("","-","?","+","0","1","2","3","4","5","6","7","8","9","B","h","H","K
","m","M")
stormDs[stormDs[,6]=="-"|stormDs[,6]=="?"|stormDs[,6]=="+"|stormDs[,6]=="",6] <- "1"
stormDs[stormDs[,6]=="h"|stormDs[,6]=="H",6] <- "2"
stormDs[stormDs[,6]=="K",6] <- "3"
stormDs[stormDs[,6]=="m"|stormDs[,6]=="M",6] <- "6"
stormDs[stormDs[,6]=="B",6] <- "9"
summary(stormDs[,6])
```

```
##              -      ?      +      0      1      2      3      4      5
##      0      0      0      0    216 465973 424684     11      4     28
##      6      7      8      9      B      h      H      K      m      M
##  11334      5      1     41      0      0      0      0      0      0
```

```
# transforming the magnitude to integer
stormDs[,6]<-as.integer(stormDs[,6])



# Check unique order levels crop damage
summary(stormDs[,8])
```

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

```
# Replace order of magnitude by numbers
levels(stormDs[,8])<-c("","-","?","+","0","1","2","3","4","5","6","7","8","9","B","h","H","K
","m","M")
stormDs[stormDs[,8]=="?"|stormDs[,6]=="",8] <- "1"
stormDs[stormDs[,8]=="h"|stormDs[,8]=="H",8] <- "2"
stormDs[stormDs[,8]=="k"|stormDs[,8]=="K",8] <- "3"
stormDs[stormDs[,8]=="m"|stormDs[,8]=="M",8] <- "6"
stormDs[stormDs[,8]=="B",8] <- "9"
summary(stormDs[,8])
```

```
##              -      ?      +      0      1      2      3      4      5
## 618413       7      0      1      9     40 281832      1   1994      0
##       6      7      8      9      B      h      H      K      m      M
##       0      0      0      0      0      0      0      0      0      0
```

```
# transforming the magnitude to integer
stormDs[,8]<-as.integer(stormDs[,8])

#Multiply magnitude by order of magnitude
stormDs[,9] <- stormDs[,5] * 10^stormDs[,6]
stormDs[,10] <- stormDs[,7] * 10^stormDs[,8]
names(stormDs)[9:10] <- c("PROPERTYDAMAGE","CROPDAMAGE")

#Subsetting
stormDs <- stormDs[,c("BGN_DATE","EVTYPE","FATALITIES","INJURIES","PROPERTYDAMAGE","CROPDAMA
GE")]
```

There are several types of storm events. For this study, these will be grouped in:

1. EROSION
2. FLOOD
3. LIGHTNING (includes thunder)
4. RAIN
5. BLIZZARD (includes snow, freeze, cold, ice)
6. HEAT
7. RAIN
8. TSUNAMI (includes tide)
9. AVALANCHE
10. VULCANIC
11. HEAT
12. TORNADO (includes hurricane, wind)

```
stormDs[grepl("EROSION",stormDs$EVTYPE),7] <- "EROSION"
stormDs[grepl("FLOOD",stormDs$EVTYPE),7] <- "FLOOD"
stormDs[grepl("LIGHTNING|THUNDER",stormDs$EVTYPE),7] <- "LIGHTNING"
stormDs[grepl("RAIN",stormDs$EVTYPE),7] <- "RAIN"
stormDs[grepl("BLIZZARD|SNOW|FREEZE|COLD|ICE",stormDs$EVTYPE),2] <- "BLIZZARD"
stormDs[grepl("HEAT",stormDs$EVTYPE),7] <- "HEAT"
stormDs[grepl("RAIN",stormDs$EVTYPE),7] <- "RAIN"
stormDs[grepl("TSUNAMI|TIDE",stormDs$EVTYPE),7] <- "TSUNAMI"
stormDs[grepl("AVALANCHE",stormDs$EVTYPE),7] <- "AVALANCHE"
stormDs[grepl("VULCANIC",stormDs$EVTYPE),7] <- "VULCANIC"
stormDs[grepl("HEAT",stormDs$EVTYPE),7] <- "HEAT"
stormDs[grepl("TORNADO|HURRICANE|WIND",stormDs$EVTYPE),7] <- "TORNADO"

names(stormDs)[7] <- "Event_type"
stormDs <- stormDs[,c("BGN_DATE","FATALITIES","INJURIES","PROPERTYDAMAGE","CROPDAMAGE","Even
t_type")]
```

To have a single measure of population health and economic consecuences, fatalities is aggregated with injuries, and property damage with crop damage:

```
stormDs[,7] <- stormDs$FATALITIES + stormDs$INJURIES
stormDs[,8] <- stormDs$PROPERTYDAMAGE + stormDs$CROPDAMAGE

names(stormDs)[7:8] <- c("Population_health","Economic_consequences")

stormDs <- stormDs[,c("BGN_DATE","Event_type","Population_health","Economic_consequences")]
```

Calculating the total amount of damages, both for Population Health and Economic Consequences

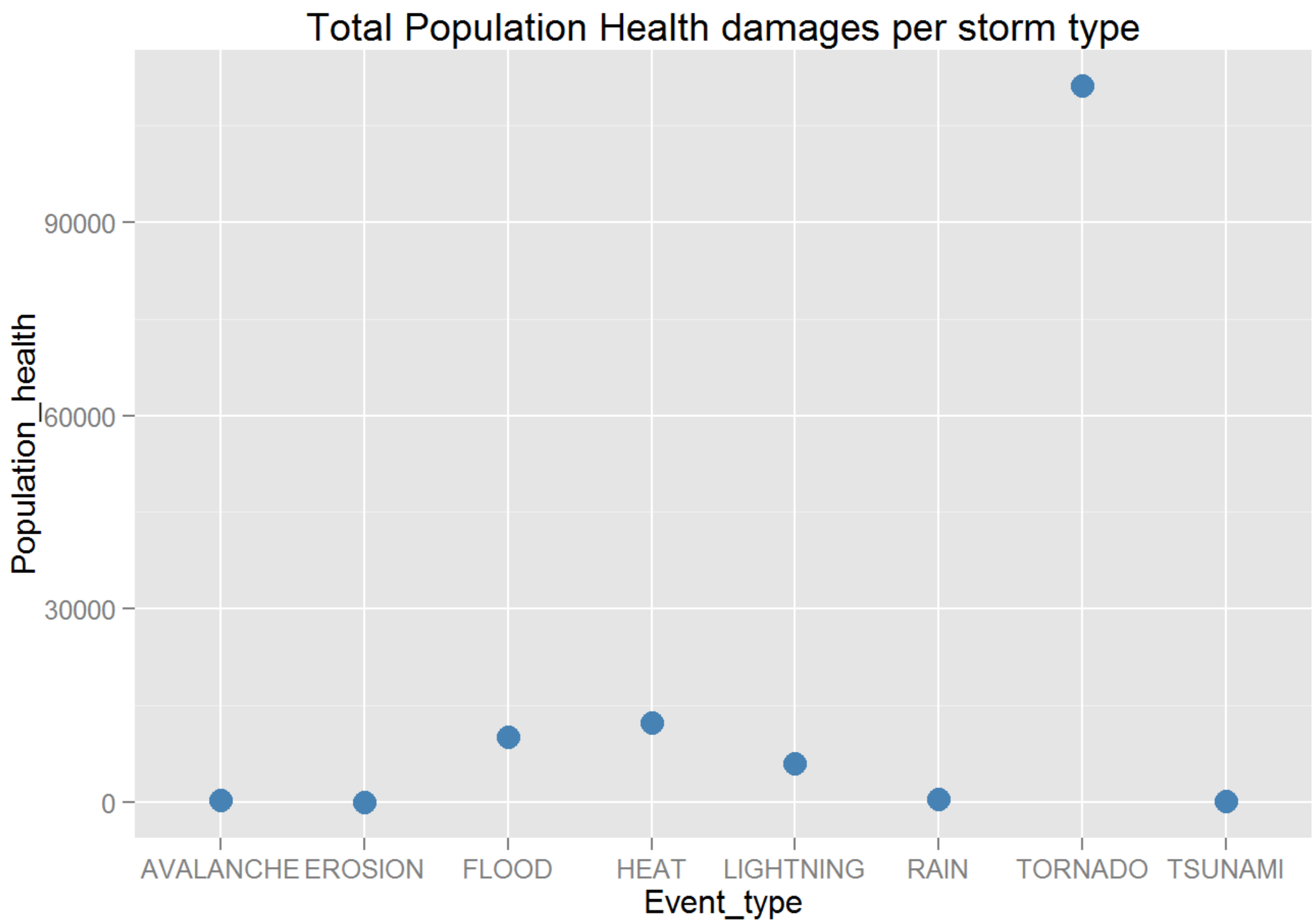
```
# Population_health
stormDsAccPH <- aggregate(Population_health ~ Event_type, stormDs, sum)

# Economic_consequences
stormDsAccEC <- aggregate(Economic_consequences ~ Event_type, stormDs, sum)
```

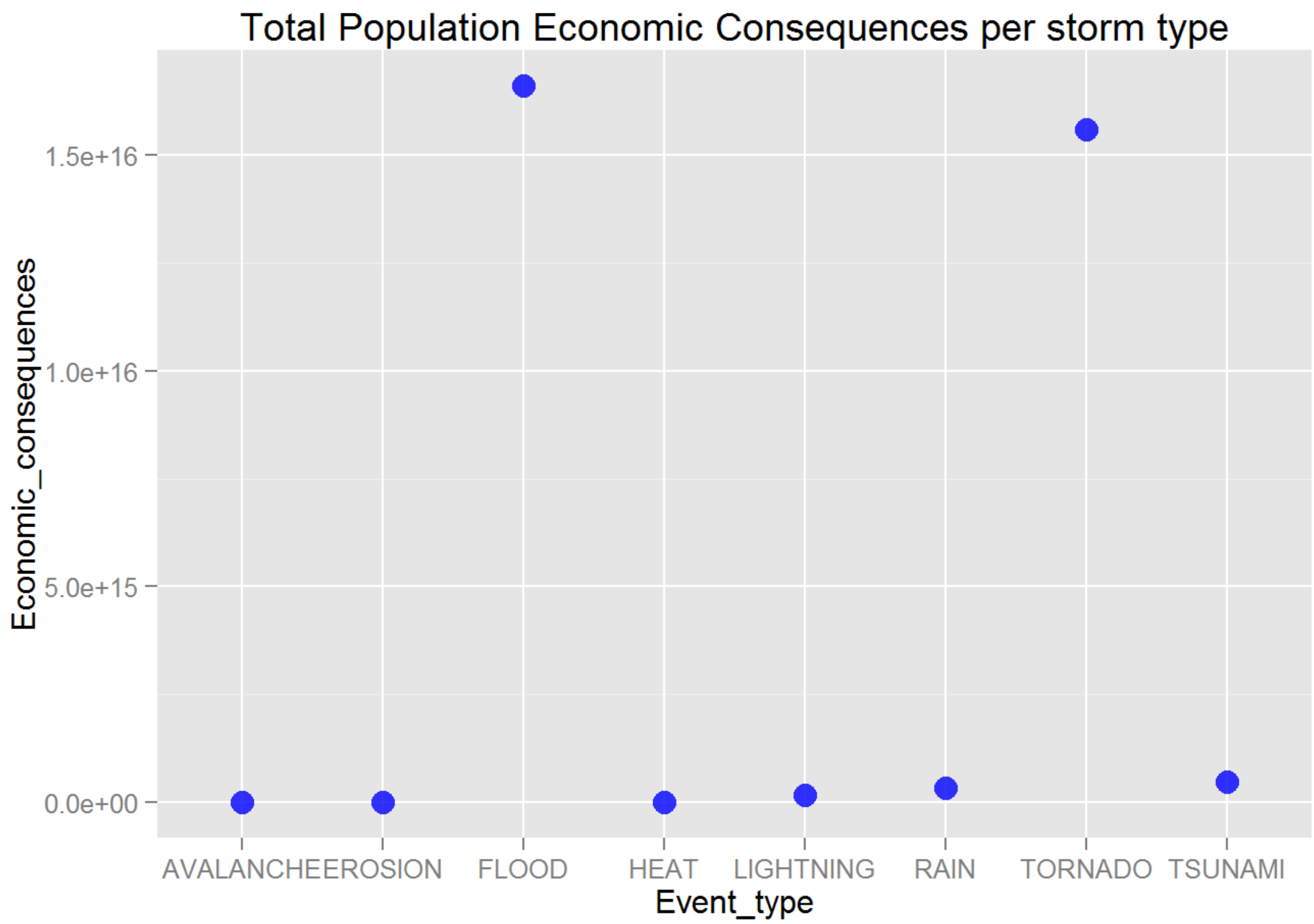
Results

To visualize better the results, here is a plot of the damages per event type:

```
# Population_health
g <- ggplot(stormDsAccPH, aes(x=Event_type,y=Population_health))
g + geom_point(color = "steelblue",size = 4, alpha = 1) + ggtitle("Total Population Health d
amages per storm type")
```




```
# Economic_consequences
g2 <- ggplot(stormDsAccEC, aes(x=Event_type,y=Economic_consequences))
g2 + geom_point(color = "blue",size = 4, alpha = 0.8) + ggtitle("Total Population Economic C
onsequences per storm type")
```



As we can see, the natural disaster that casuses the most Population Health damages are tornados, followed by heat and floods. On the other hand, the events that cause the most economic losses are floods followed by tornados.