Storm Data

Diana Marcela Rosero

26/7/2020

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

Synonpsis

In this report, we aim to analyze the impact of different weather events on public health and economy based on the storm database collected from the U.S. National Oceanic and Atmospheric Administration's (NOAA) from 1950 - 2011. We will use the estimates of fatalities, injuries, property and crop damage to decide which types of event are most harmful to the population health and economy. From these data, we found that excessive heat and tornado are most harmful with respect to population health, while flood, drought, and hurricane/typhoon have the greatest economic consequences.

Basic settings

```r
# Import data
setwd("E:/ESCRITORIO/Desktop/specdata")
storm.data <- read.csv("repdata_data_StormData.csv")
# Libraries
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 4.0.2
```

```
##
## 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
```

```r
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.0.2
```

```r
library(plyr)
```

```
## Warning: package 'plyr' was built under R version 4.0.2
```

```
## ------------------------------------------------------------------------------
```

```
## You have loaded plyr after dplyr - this is likely to cause problems.
## If you need functions from both plyr and dplyr, please load plyr first, then dplyr:
## library(plyr); library(dplyr)
## ------------------------------------------------------------------------------

##
## Attaching package: 'plyr'

## The following objects are masked from 'package:dplyr':
##
##     arrange, count, desc, failwith, id, mutate, rename, summarise,
##     summarize
```

```
Sys.setlocale("LC_TIME", "English")
```

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

```
# some information about the variables
str(storm.data)
```

```
## 'data.frame':    902297 obs. of  37 variables:
##  $ STATE__   : num  1 1 1 1 1 1 1 1 1 1 ...
##  $ BGN_DATE  : chr  "4/18/1950 0:00:00" "4/18/1950 0:00:00" "2/20/1951 0:00:00" "6/8/1951 0:00:00" .
##  $ BGN_TIME  : chr  "0130" "0145" "1600" "0900" ...
##  $ TIME_ZONE : chr  "CST" "CST" "CST" "CST" ...
##  $ COUNTY    : num  97 3 57 89 43 77 9 123 125 57 ...
##  $ COUNTYNAME: chr  "MOBILE" "BALDWIN" "FAYETTE" "MADISON" ...
##  $ STATE     : chr  "AL" "AL" "AL" "AL" ...
##  $ EVTYPE    : chr  "TORNADO" "TORNADO" "TORNADO" "TORNADO" ...
##  $ BGN_RANGE : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ BGN_AZI   : chr  "" "" "" "" ...
##  $ BGN_LOCATI: chr  "" "" "" "" ...
##  $ END_DATE  : chr  "" "" "" "" ...
##  $ END_TIME  : chr  "" "" "" "" ...
##  $ 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   : chr  "" "" "" "" ...
##  $ END_LOCATI: chr  "" "" "" "" ...
##  $ 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: chr  "K" "K" "K" "K" ...
##  $ CROPDMG   : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ CROPDMGEXP: chr  "" "" "" "" ...
##  $ WFO       : chr  "" "" "" "" ...
##  $ STATEOFFIC: chr  "" "" "" "" ...
##  $ ZONENAMES : chr  "" "" "" "" ...
##  $ 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   : chr  "" "" "" "" ...
##  $ REFNUM    : num  1 2 3 4 5 6 7 8 9 10 ...
```

```
summary(storm.data)
```

```
##     STATE__       BGN_DATE           BGN_TIME          TIME_ZONE
##  Min.   : 1.0   Length:902297      Length:902297      Length:902297
##  1st Qu.:19.0   Class :character   Class :character   Class :character
##  Median :30.0   Mode  :character   Mode  :character   Mode  :character
##  Mean   :31.2
##  3rd Qu.:45.0
##  Max.   :95.0
##
##      COUNTY       COUNTYNAME           STATE              EVTYPE
##  Min.   :  0.0   Length:902297      Length:902297      Length:902297
##  1st Qu.: 31.0   Class :character   Class :character   Class :character
##  Median : 75.0   Mode  :character   Mode  :character   Mode  :character
##  Mean   :100.6
##  3rd Qu.:131.0
##  Max.   :873.0
##
##    BGN_RANGE          BGN_AZI           BGN_LOCATI          END_DATE
##  Min.   :   0.000   Length:902297      Length:902297      Length:902297
##  1st Qu.:   0.000   Class :character   Class :character   Class :character
##  Median :   0.000   Mode  :character   Mode  :character   Mode  :character
##  Mean   :   1.484
##  3rd Qu.:   1.000
##  Max.   :3749.000
##
##    END_TIME           COUNTY_END COUNTYENDN       END_RANGE
##  Length:902297      Min.   :0    Mode:logical   Min.   :  0.0000
##  Class :character   1st Qu.:0    NA's:902297    1st Qu.:  0.0000
##  Mode  :character   Median :0                   Median :  0.0000
##                     Mean   :0                   Mean   :  0.9862
##                     3rd Qu.:0                   3rd Qu.:  0.0000
##                     Max.   :0                   Max.   :925.0000
##
##    END_AZI           END_LOCATI           LENGTH             WIDTH
##  Length:902297      Length:902297      Min.   :   0.0000   Min.   :   0.000
##  Class :character   Class :character   1st Qu.:   0.0000   1st Qu.:   0.000
##  Mode  :character   Mode  :character   Median :   0.0000   Median :   0.000
##                                        Mean   :   0.2301   Mean   :   7.503
##                                        3rd Qu.:   0.0000   3rd Qu.:   0.000
##                                        Max.   :2315.0000   Max.   :4400.000
##
##        F               MAG            FATALITIES          INJURIES
##  Min.   :0.0      Min.   :    0.0   Min.   :  0.0000   Min.   :   0.0000
##  1st Qu.:0.0      1st Qu.:    0.0   1st Qu.:  0.0000   1st Qu.:   0.0000
##  Median :1.0      Median :   50.0   Median :  0.0000   Median :   0.0000
##  Mean   :0.9      Mean   :   46.9   Mean   :  0.0168   Mean   :   0.1557
##  3rd Qu.:1.0      3rd Qu.:   75.0   3rd Qu.:  0.0000   3rd Qu.:   0.0000
##  Max.   :5.0      Max.   :22000.0   Max.   :583.0000   Max.   :1700.0000
##  NA's   :843563
##     PROPDMG          PROPDMGEXP           CROPDMG          CROPDMGEXP
##  Min.   :   0.00   Length:902297      Min.   :  0.000   Length:902297
##  1st Qu.:   0.00   Class :character   1st Qu.:  0.000   Class :character
##  Median :   0.00   Mode  :character   Median :  0.000   Mode  :character
##  Mean   :  12.06                      Mean   :  1.527
##  3rd Qu.:   0.50                      3rd Qu.:  0.000
##  Max.   :5000.00                      Max.   :990.000
##
##      WFO             STATEOFFIC         ZONENAMES           LATITUDE
##  Length:902297      Length:902297      Length:902297      Min.   :   0
##  Class :character   Class :character   Class :character   1st Qu.:2802
##  Mode  :character   Mode  :character   Mode  :character   Median :3540
##                                                           Mean   :2875
##                                                           3rd Qu.:4019
##                                                           Max.   :9706
##                                                           NA's   :47
##    LONGITUDE        LATITUDE_E     LONGITUDE_       REMARKS
##  Min.   :-14451   Min.   :   0   Min.   :-14455   Length:902297
##  1st Qu.:  7247   1st Qu.:   0   1st Qu.:     0   Class :character
##  Median :  8707   Median :   0   Median :     0   Mode  :character
##  Mean   :  6940   Mean   :1452   Mean   :  3509
##  3rd Qu.:  9605   3rd Qu.:3549   3rd Qu.:  8735
##  Max.   : 17124   Max.   :9706   Max.   :106220
##                   NA's   :40
##      REFNUM
##  Min.   :     1
##  1st Qu.:225575
##  Median :451149
##  Mean   :451149
##  3rd Qu.:676723
##  Max.   :902297
##
```

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

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

There are 902297 rows and 37 columns in total. The events in the database start in the year 1950 and end in November 2011. In the earlier years of the database there are generally fewer events recorded, most likely due to a lack of good records. More recent years should be considered more complete.

Extracting variables of interest for analysis of weather impact on health and economy

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

```
##                 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
```

Checking for missing values

```
sum(is.na(mydata$FATALITIES))
```

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

```
sum(is.na(mydata$INJURIES))
```

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

```
# Check for missing values in economic variables for "size" of damage - there is no NA's in the data.
sum(is.na(mydata$PROPDMG))
```

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

```
sum(is.na(mydata$CROPDMG))
```

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

```
# Check for missing values in economic variables for units damage - there is no NA's in the data.
sum(is.na(mydata$PROPDMGEXP))
```

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

```
sum(is.na(mydata$CROPDMGEXP))
```

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

Transforming extracted variables

```
sort(table(mydata$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
```

```
# create a new variable EVENT to transform variable EVTYPE in groups
mydata$EVENT <- "OTHER"
# group by keyword in EVTYPE
mydata$EVENT[grep("HAIL", mydata$EVTYPE, ignore.case = TRUE)] <- "HAIL"
mydata$EVENT[grep("HEAT", mydata$EVTYPE, ignore.case = TRUE)] <- "HEAT"
mydata$EVENT[grep("FLOOD", mydata$EVTYPE, ignore.case = TRUE)] <- "FLOOD"
mydata$EVENT[grep("WIND", mydata$EVTYPE, ignore.case = TRUE)] <- "WIND"
mydata$EVENT[grep("STORM", mydata$EVTYPE, ignore.case = TRUE)] <- "STORM"
mydata$EVENT[grep("SNOW", mydata$EVTYPE, ignore.case = TRUE)] <- "SNOW"
mydata$EVENT[grep("TORNADO", mydata$EVTYPE, ignore.case = TRUE)] <- "TORNADO"
mydata$EVENT[grep("WINTER", mydata$EVTYPE, ignore.case = TRUE)] <- "WINTER"
mydata$EVENT[grep("RAIN", mydata$EVTYPE, ignore.case = TRUE)] <- "RAIN"
# listing the transformed event types
sort(table(mydata$EVENT), decreasing = TRUE)
```

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

```
# Checking the values for variables that represent units od dollars
sort(table(mydata$PROPDMGEXP), decreasing = TRUE)[1:10]
```

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

```
sort(table(mydata$CROPDMGEXP), decreasing = TRUE)[1:10]
```

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

There is some mess in units, so we transform those variables in one unit (dollar) variable by the following rule: * 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

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

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

```
# Print of first 10 values for property damage (in dollars) that most appear in the data
sort(table(mydata$property.damage), decreasing = TRUE)[1:10]
```

```
##
##      0   5000  10000   1000   2000  25000  50000   3000  20000  15000
## 663123  31731  21787  17544  17186  17104  13596  10364   9179   8617
```

```
# Print of first 10 values for crop damage (in dollars) that most appear in the data
sort(table(mydata$crop.damage), decreasing = TRUE)[1:10]
```

```
##
##      0   5000  10000  50000  1e+05   1000   2000  25000  20000  5e+05
## 880198   4097   2349   1984   1233    956    951    830    758    721
```

Analysis

Aggregating events for public health variables

Table of public health problems by event type

```
# aggregate FATALITIES and INJURIES by type of EVENT
agg.fatalites.and.injuries <- ddply(mydata, .(EVENT), summarize, Total = sum(FATALITIES + INJURIES,  na
agg.fatalites.and.injuries$type <- "fatalities and injuries"

# aggregate FATALITIES by type of EVENT
agg.fatalities <- ddply(mydata, .(EVENT), summarize, Total = sum(FATALITIES, na.rm = TRUE))
agg.fatalities$type <- "fatalities"

# aggregate INJURIES by type of EVENT
agg.injuries <- ddply(mydata, .(EVENT), summarize, Total = sum(INJURIES, na.rm = TRUE))
agg.injuries$type <- "injuries"

# combine all
agg.health <- rbind(agg.fatalities, agg.injuries)

health.by.event <- join (agg.fatalities, agg.injuries, by="EVENT", type="inner")
health.by.event
```

```
##      EVENT Total       type Total     type
## 1    FLOOD  1524 fatalities  8602 injuries
## 2     HAIL    15 fatalities  1371 injuries
## 3     HEAT  3138 fatalities  9224 injuries
## 4    OTHER  2626 fatalities 12224 injuries
## 5     RAIN   114 fatalities   305 injuries
## 6     SNOW   164 fatalities  1164 injuries
## 7    STORM   416 fatalities  5339 injuries
## 8  TORNADO  5661 fatalities 91407 injuries
## 9     WIND  1209 fatalities  9001 injuries
## 10  WINTER   278 fatalities  1891 injuries
```

Aggregating events for economic variables

```
# aggregate PropDamage and CropDamage by type of EVENT
agg.propdmg.and.cropdmg <- ddply(mydata, .(EVENT), summarize, Total = sum(property.damage + crop.damage
agg.propdmg.and.cropdmg$type <- "property and crop damage"

# aggregate PropDamage by type of EVENT
agg.prop <- ddply(mydata, .(EVENT), summarize, Total = sum(property.damage, na.rm = TRUE))
agg.prop$type <- "property"

# aggregate INJURIES by type of EVENT
agg.crop <- ddply(mydata, .(EVENT), summarize, Total = sum(crop.damage, na.rm = TRUE))
agg.crop$type <- "crop"

# combine all
agg.economic <- rbind(agg.prop, agg.crop)


economic.by.event <- join (agg.prop, agg.crop, by="EVENT", type="inner")
economic.by.event
```

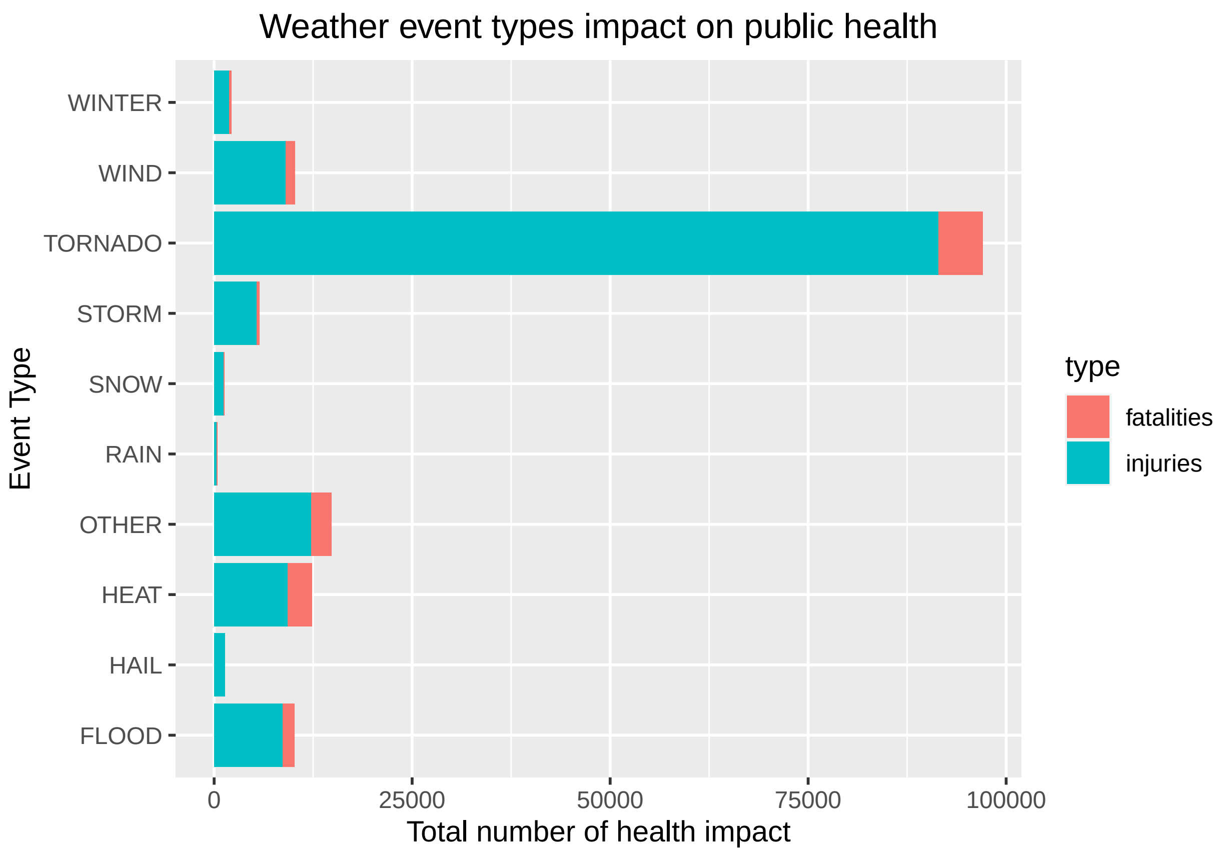
```
##      EVENT        Total     type       Total type
## 1    FLOOD 167502193929 property 12266906100 crop
## 2     HAIL  15733043048 property  3046837473 crop
## 3     HEAT     20325750 property   904469280 crop
## 4    OTHER  97246712337 property 23588880870 crop
## 5     RAIN   3270230192 property   919315800 crop
## 6     SNOW   1024169752 property   134683100 crop
## 7    STORM  66304415393 property  6374474888 crop
## 8  TORNADO  58593098029 property   417461520 crop
## 9     WIND  10847166618 property  1403719150 crop
## 10  WINTER   6777295251 property    47444000 crop
```

Results

Across the United States, which types of events are most harmful with respect to population health?

```
# transform EVENT to factor variable for health variables
agg.health$EVENT <- as.factor(agg.health$EVENT)

# plot FATALITIES and INJURIES by EVENT
health.plot <- ggplot(agg.health, aes(x = EVENT, y = Total, fill = type)) + geom_bar(stat = "identity")
  coord_flip() +
  xlab("Event Type") +
  ylab("Total number of health impact") +
  ggtitle("Weather event types impact on public health") +
  theme(plot.title = element_text(hjust = 0.5))
print(health.plot)
```

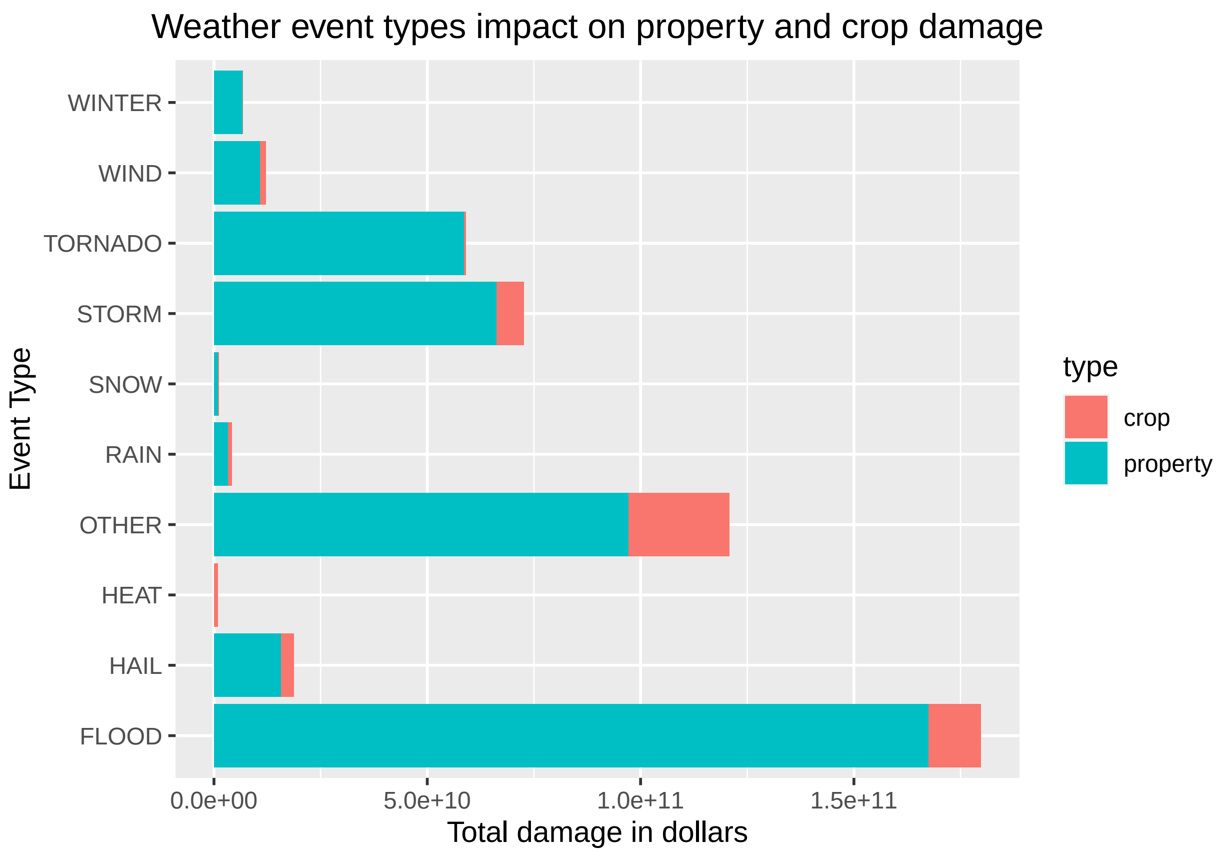


The most harmful weather event for health (in number of total fatalites and injuries) is, by far, a tornado.

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

```
#transform EVENT to factor variable for economic variables
agg.economic$EVENT <- as.factor(agg.economic$EVENT)

# plot PROPERTY damage and CROP damage by EVENT
economic.plot <- ggplot(agg.economic, aes(x = EVENT, y = Total, fill = type)) + geom_bar(stat = "identi
  coord_flip() +
  xlab("Event Type") +
  ylab("Total damage in dollars") +
  ggtitle("Weather event types impact on property and crop damage") +
  theme(plot.title = element_text(hjust = 0.5))
print(economic.plot)
```



The most devastating weather event with the greatest economic cosequences (to property and crops) is a flood.

Conclusion

From these data, we found that **excessive heat** and **tornado** are most harmful with respect to population health, while **flood**, **drought**, and **hurricane/typhoon** have the greatest economic consequences.