Impact of meteorological events on population and economy in the US

FK

April 2021

Synposis: This project is based on the *U.S. National Oceanic and Atmospheric Administration's (NOAA) storm database*, which tracks characteristics of major storms and weather events in the United States. We use the *Storms dataset* to study the impact of meteorological events on the population health and on the economy. Our analysis shows that storms and extreme heat are the most harmful event types with respect to population health. Fires as well as events related to extreme temperatures and precipitation are the most relevant event types with regard to the economic consequences; these are measured by the damage of property and the damage of crop, which the events caused.

Data Processing

First of all, we download the *Storms Dataset* and import it using read.csv.

```
fileUrl <- "https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2"

if(!file.exists("StormsData.csv.bz2")){
  download.file(fileUrl,"StormsData.csv.bz2")
}
data <- read.csv("StormsData.csv.bz2")
```

For better handling, we format all column names to lower cases letters.

```
colnames(data) <- tolower(colnames(data))
```

In the *Storms database*, alphabetical characters are used to signify magnitude of damage, both for property damage and crop damage.

```
table(data$propdmgexp)
```

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

According to the *Storm Dataset* documentation, we have the following correspondences:

- "H"/"h" means hundreds,
- "K"/"k" means thousands,
- "M"/"m" means millions,
- "B"/"b" means billions.

It is not clear what the letters '+','?' and '-' stand for.

We create new columns for each damage type (i.e., "property damage", "crop damage") which state the damage in US Dollars as a total number. This is going to be more suited for the data analysis, which we will perform below. For the letters '+','-','?' for which we do not know what they signify we set the damage to be 'NA'. This is not too bad because these letters are very rarely used, see the above table.

```
calculateDamage <- function(num,exponent){
  if(exponent=="H" | exponent=="h"){
    num*10^2
  }
  else if(exponent=="K" | exponent=="k"){
    num*10^3
  }
  else if(exponent=="M" | exponent=="m"){
    num*10^6
  }
  else if(exponent=="B" | exponent=="b"){
    num*10^9
  }
  else if(exponent %in% c(0:10)){
    num*10^as.integer(exponent)
  }
  else if(exponent %in% c('+','-','?')){
    return(NA)
  }
  else{
    0
  }
}
data$property_damage = mapply(calculateDamage,data$propdmg,data$propdmgexp)
data$crop_damage = mapply(calculateDamage,data$cropdmg,data$cropdmgexp)
```

The *Storms Dataset* considers a large number of types of events:

```
length(unique(data$evtype))
```

```
## [1] 985
```

If we look at the table of event types, then we see that some of them differ only in spelling (e.g. FLOOD vs. FLOODS) or in lower/upper case letters (e.g. HEAVY PRECIPITATION vs. Heavy Precipitation). Firstly, we convert everything to lower case letters and remove superfluous whitespaces

```
data$evtype <- tolower(data$evtype)
data$evtype <- str_trim(data$evtype,side="both")
```

For record:

```
length(unique(data$evtype))
```

```
## [1] 890
```

As can be seen, the number of event types has already significantly reduced. The most common ones (used more than 100 times) are:

```
tmp<- table(data$evtype)
sort(tmp[which(tmp>100)],decreasing=TRUE)
```

```
##
##                    hail              tstm wind       thunderstorm wind
##                  288661                 219946                   82564
##                 tornado            flash flood                   flood
##                   60652                  54278                   25327
```

```
##       thunderstorm winds                  high wind                  lightning
##                    20843                      20214                      15755
##               heavy snow                 heavy rain               winter storm
##                    15708                      11742                      11433
##           winter weather               funnel cloud           marine tstm wind
##                     7045                       6844                       6175
## marine thunderstorm wind                 waterspout                strong wind
##                     5812                       3797                       3569
##     urban/sml stream fld                   wildfire                   blizzard
##                     3392                       2761                       2719
##                  drought                  ice storm             excessive heat
##                     2488                       2006                       1678
##               high winds           wild/forest fire               frost/freeze
##                     1533                       1457                       1343
##                dense fog         winter weather/mix             tstm wind/hail
##                     1293                       1104                       1028
##  extreme cold/wind chill                       heat                  high surf
##                     1002                        767                        734
##           tropical storm             flash flooding              coastal flood
##                      690                        682                        657
##             extreme cold           lake-effect snow          flood/flash flood
##                      657                        636                        625
##                     snow                  landslide            cold/wind chill
##                      617                        600                        539
##                      fog                rip current                marine hail
##                      538                        470                        442
##               dust storm                  avalanche                       wind
##                      427                        386                        347
##             rip currents                storm surge              freezing rain
##                      304                        261                        260
##              urban flood       heavy surf/high surf          extreme windchill
##                      251                        228                        204
##             strong winds             dry microburst           coastal flooding
##                      204                        186                        183
##               light snow      astronomical low tide                  hurricane
##                      176                        174                        174
##              river flood              record warmth                 dust devil
##                      173                        154                        149
##         storm surge/tide           marine high wind          unseasonably warm
##                      148                        135                        126
##                 flooding     astronomical high tide          moderate snowfall
##                      120                        103                        101
```

We make the following replacements to take care of differences in spelling.

```
data$evtype <- gsub(" /|/ |/","-",data$evtype)
data$evtype <- gsub("\\s\\s"," ",data$evtype)
data$evtype <- gsub(" and |\\s?\\&\\s?|\\s?;\\s?","-",data$evtype)
data$evtype <- gsub("\\.$|-$","",data$evtype)
data$evtype <- gsub("fires","fire",data$evtype)
data$evtype <- gsub("(mud|land)\\s?slides?","mud slide",data$evtype)
data$evtype <- gsub("currents","current",data$evtype)
data$evtype <- gsub("snowfall","snow",data$evtype)
data$evtype <- gsub("floods|fld$|flooding","flood",data$evtype)
data$evtype <- gsub("wa(y)?ter(\\s)?spout(s)?","waterspout",data$evtype)
```

```
data$evtype <- gsub("winds|windss|wins|wnd$","wind",data$evtype)
data$evtype <- gsub("rains|rainfall","rain",data$evtype)
data$evtype <- gsub("wint(e)?r(y)?\\s?(weather)?-?\\s?(mix)?","wintery weather mix",data$evtype)
data$evtype <- gsub("tides","tide",data$evtype)
data$evtype <- gsub("tstm|thunderstorms","thunderstorm",data$evtype)
data$evtype <- gsub("thunderstormw","thunderstorm wind",data$evtype)
data$evtype <- gsub("torndao|tornados","tornado",data$evtype)
data$evtype <- gsub("unseasonal|unusual|unusally","unseasonably",data$evtype)
data$evtype <- gsub("^(sml)","small",data$evtype)
data$evtype <- gsub("(hvy)","heavy",data$evtype)
data$evtype <- gsub("strm","stream",data$evtype)
data$evtype <- gsub("rain \\(heavy\\)","heavy rain",data$evtype)
data$evtype <- gsub("seas$","sea",data$evtype)
data$evtype <- gsub("temperatures","temperature",data$evtype)
data$evtype <- gsub("avalance","avalanche",data$evtype)
data$evtype <- gsub("coastalflood","coastal flood",data$evtype)
data$evtype <- gsub("icy roads","ice roads",data$evtype)
data$evtype <- gsub("record(\\s)?","",data$evtype)
```

For record:

```
length(unique(data$evtype))
```

```
## [1] 737
```

We have focussed here on the most important changes since the analysis which we are going to do does not go into details. If one wants to perform a deep analysis, one should have a closer look at the names of the event types and unify them further. We will group the events into some larger classes:

- events related to cold/hot temperature
- events related to wind (storm, typhoon, funnel cloud, hurricane)
- events related to precipitation (drought, rain, hail, snow,fog)
- events related to floods
- events related to fires
- events related to avalanches/slides

We create a new column for the event class.

```
keywordsTemp <- c("cold","heat","hot","warm","freeze","frost")
keywordsWind <- c("wind","storm","typhoon","hurricane","funnel cloud","dust","tornado")
keywordsPrec <- c("drought","dry","rain","hail","snow","fog","sleet","ice","lightning","waterspout")
keywordsFire <- c("fire")
keywordsFlood <- c("flood","surf","current","tide")
keywordsAval <- c("avalanche","slide")

data<- mutate(data,eventclass=case_when(str_detect(data$evtype,paste(keywordsTemp,collapse="|"))~"tempe
  str_detect(data$evtype,paste(keywordsWind,collapse="|"))~"wind",
  str_detect(data$evtype,paste(keywordsPrec,collapse="|"))~"precipitation",
  str_detect(data$evtype, paste(keywordsFire,collapse="|"))~"fire",
  str_detect(data$evtype, paste(keywordsFlood,collapse="|"))~"flood",
  str_detect(data$evtype, paste(keywordsAval,collapse="|"))~"avalanche"))

data$eventclass[is.na(data$eventclass)]<-"other"

table(data$eventclass)
```

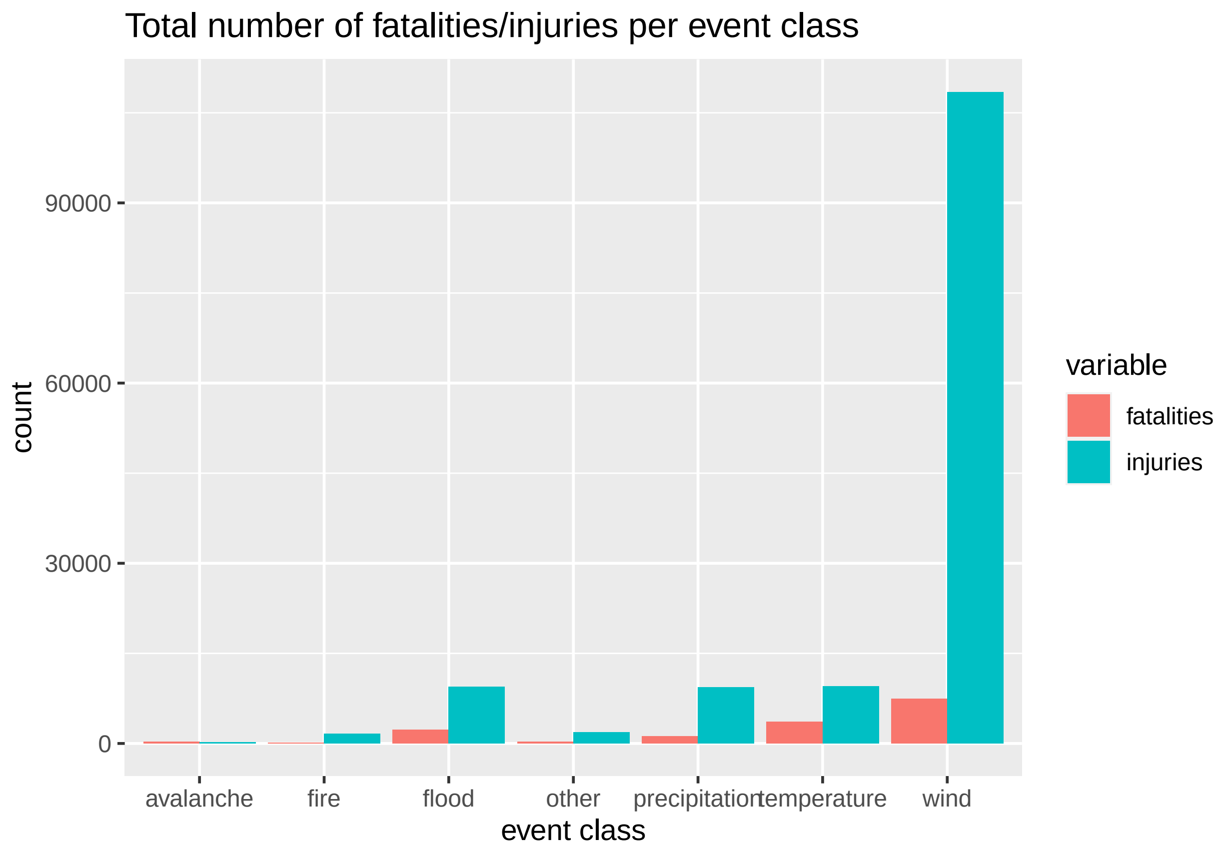
```
##
```

```
##     avalanche          fire         flood         other precipitation
##          1033          4239         88186         11762        343543
##   temperature          wind
##          6959        446575
```

Data Analysis

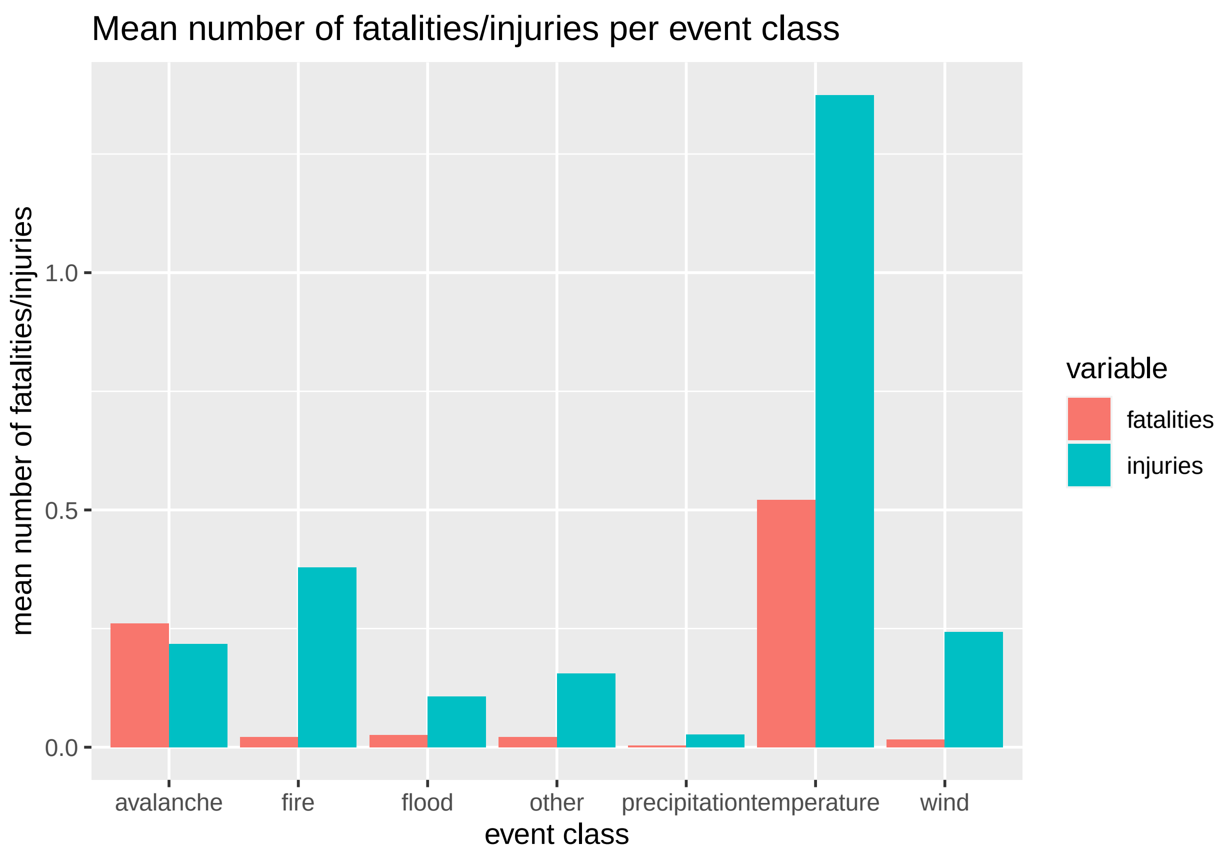
In this section, we study the data which we just prepared for analysis. The first question which we want to study is which types of events are the most harmful with respect to population health. We take a look at the total number of fatalities/injuries for each event class:

```
totalByClass <- aggregate(list(fatalities=data$fatalities,injuries=data$injuries),list(class=data$eventc

totalByClass$ID <- as.factor(1:nrow(totalByClass))
ggplot(melt(totalByClass, id=c("ID","class"),value.name = "count"),aes(class,count, fill=variable, grou
     geom_bar(stat='identity', position='dodge')+
     labs(y="count",x="event class",title="Total number of fatalities/injuries per event class")
```



Each event class contains a much different number of events, and so we cannot compare directly the total number of fatalities. It is better to consider the mean fatalities for each event class:

```
meanByClass <- aggregate(list(fatalities=data$fatalities,injuries=data$injuries),list(class=data$eventcl

meanByClass$ID <- as.factor(1:nrow(meanByClass))
ggplot(melt(meanByClass, id=c("ID","class"),value.name = "mean"),aes(class,mean, fill=variable, group=va
     geom_bar(stat='identity', position='dodge')+
     labs(y="mean number of fatalities/injuries",x="event class",title="Mean number of fatalities/injur
```



There are two classes which are of particular interest: wind (because of the high total number of fatalities/injuries) and temperature (high mean number of fatalities/injuries). The following tables show for the classes "wind" and "temperature" the mean number of fatalities and injuries per event type:

```
mapply(function(str){round(mean(filter(data,grepl(str,data$evtype))$fatalities,rm.na=TRUE),digits=2)},k
```

```
##          wind         storm       typhoon     hurricane funnel cloud          dust
##          0.00          0.00          0.65          0.47          0.00          0.04
##       tornado
##          0.09
```

```
mapply(function(str){round(mean(filter(data,grepl(str,data$evtype))$injuries,rm.na=TRUE),digits=2)},key
```

```
##          wind         storm       typhoon     hurricane funnel cloud          dust
##          0.03          0.04         12.93          4.61          0.00          0.82
##       tornado
##          1.51
```

```
mapply(function(str){round(mean(filter(data,grepl(str,data$evtype))$fatalities,rm.na=TRUE),digits=2)},k
```

```
##   cold   heat    hot   warm freeze  frost
##   0.18   1.19   0.00   0.12   0.00   0.00
```

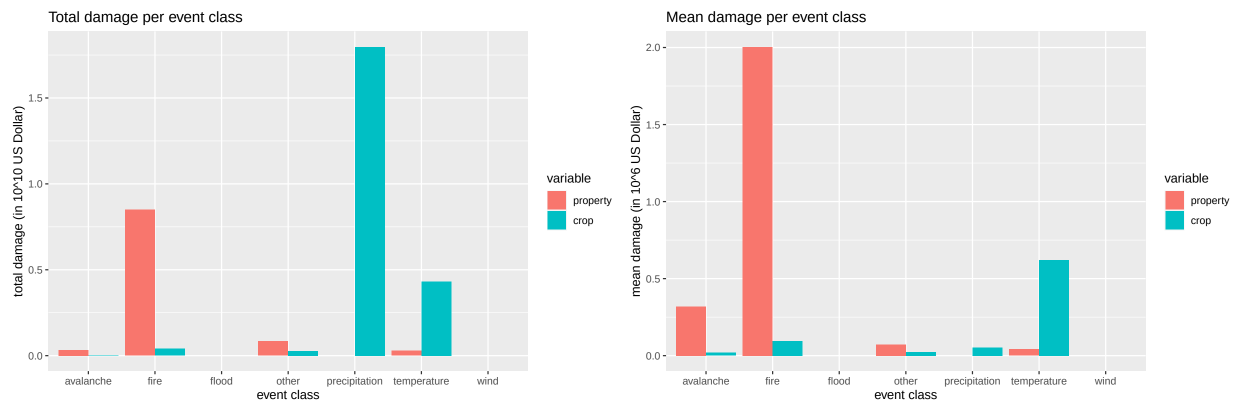
```
mapply(function(str){round(mean(filter(data,grepl(str,data$evtype))$injuries,rm.na=TRUE),digits=2)},key
```

```
##   cold   heat    hot   warm freeze  frost
##   0.13   3.48   0.00   0.06   0.00   0.00
```

To summarize: We have seen that events related to temperature and wind are very harmful with respect to population health. Particularly harmful are heat as well as heavy storms. The total number of injuries/fatalities caused by avalanches or fires are relatively small, but the mean number is relatively high, i.e. whenever the a fire or avalanche occurs the number of victims is quite large.

Next we address the greatest economic consequences. We use the same event classes as before and calculate for each event class the property damage and the crop damage.

```
damageByClass <- aggregate(list(property=data$property_damage,crop=data$crop_damage),list(class=data$eve
MeanDamageByClass <- aggregate(list(property=data$property_damage,crop=data$crop_damage),list(class=data

damageByClass$ID <- as.factor(1:nrow(damageByClass))
MeanDamageByClass$ID <- as.factor(1:nrow(MeanDamageByClass))

p1<- ggplot(melt(damageByClass, id=c("ID","class"),value.name = "damage"),aes(class,damage/10^10, fill=
     geom_bar(stat='identity', position='dodge')+
     labs(x="event class",y="total damage (in 10^10 US Dollar)")+
     labs(title="Total damage per event class")
p2 <- ggplot(melt(MeanDamageByClass, id=c("ID","class"),value.name = "damage"),aes(class,damage/10^6, fi
     geom_bar(stat='identity', position='dodge')+
     labs(x="event class",y="mean damage (in 10^6 US Dollar)")+
     labs(title="Mean damage per event class")
p1+p2
```



We see that fires cause the highest property damage (both in total and mean). For the crop damage, there are two important classes of events: those related to precipitation (e.g. droughts) and related to temperature (e.g. heat waves) have a big impact.

Results

The largest number of events in the *Storms dataset* is related to precipitation and winds.

```
table(data$eventclass)
```

```
##
##     avalanche          fire         flood         other precipitation
##          1033          4239         88186         11762        343543
##   temperature          wind
##          6959        446575
```

Our data analysis shows that the event types "storms" (including typhoons, hurricanes,. . .) and "heat" are the most harmful ones with respect to population health. Avalanches and fires happen less frequently than, say, e.g. storms, and so their total impact on population health is not that big. However, the mean number of injuries is for both event classes relatively high. Regarding the economic impact, one has to distinguish between damage of property and damage of crops. Fires are the most harmful events with respect to damage of property, while events related to precipitation and temperature (e.g. droughts, heat waves) cause a large part of the crop damage.

Discussion

It would be interesting to see how the damages/fatalities/injuries have evolved over time. The *Storms dataset* is somewhat biased because for the earlier years of the database there are generally fewer events recorded. Because of the incompleteness of the records for the earlier years, one might consider studying only the events in recent time.