Analysis of severe weather events impact on population health and the economy across the United States

ourbanow

5/20/2020

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. The present analysis summarizes and ranks the impact of severe weather events across the United States in terms of fatalities, injuries and property damage, per type of event.

Data Processing

The data used for this analysis comes from the U.S. National Oceanic and Atmospheric Administration's (NOAA) storm database.

```
fileURL <- "https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2"
if (!file.exists("Storm_Data.csv.bz2")) {
    download.file(fileURL, "Storm_Data.csv.bz2")
}
Storm_data <- read.csv("StormData.csv")
```

A copy of the data can be downloaded from here: https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2
Additional information about this data is available here:
* National Weather Service Storm Data Dcoumentation
* National Climatic Data Center Storm Events FAQ

The database contains a total of 902297 events, for which 37 variables where measured.

```
names(Storm_data)
```

```
##  [1] "STATE__"    "BGN_DATE"   "BGN_TIME"   "TIME_ZONE"  "COUNTY"
##  [6] "COUNTYNAME" "STATE"      "EVTYPE"     "BGN_RANGE"  "BGN_AZI"
## [11] "BGN_LOCATI" "END_DATE"   "END_TIME"   "COUNTY_END" "COUNTYENDN"
## [16] "END_RANGE"  "END_AZI"    "END_LOCATI" "LENGTH"     "WIDTH"
## [21] "F"          "MAG"        "FATALITIES" "INJURIES"   "PROPDMG"
## [26] "PROPDMGEXP" "CROPDMG"    "CROPDMGEXP" "WFO"        "STATEOFFIC"
## [31] "ZONENAMES"  "LATITUDE"   "LONGITUDE"  "LATITUDE_E" "LONGITUDE_"
## [36] "REMARKS"    "REFNUM"
```

For this analysis, we are only focusing on the impact of severe weather event, per type, on population health and the economy. For further processing we will then ignore non-relevant variables.

Specifically, according to NOAA,
* The data recording start from Jan. 1950. At that time they recorded one event type, tornado. They add more events gradually and only from Jan. 1996 they start recording all events type. Since our objective is comparing the effects of different weather events, we will focus only on event from Jan. 1996 on.

```
library(dplyr)
stormdata <- Storm_data[,c("BGN_DATE", "EVTYPE", "FATALITIES", "INJURIES", "PROPDMG", "PROPDMGEXP", "CR
stormdata$BGN_DATE <- as.Date(as.character(stormdata$BGN_DATE), "%m/%d/%Y %H:%M:%S")
stormdata <- filter(stormdata, BGN_DATE >= as.Date("1996-01-01"))
```

- Property and Crop damages are expressed in dollars, using different scales which we need to reprocess with the following reading keys:
  - B or b = Billion, M or m = Million, K or k = Thousand, H or h = Hundred
  - The number from one to ten represent the power of ten (10^The number)
  - The symbols "-", "+" and "?" refer to less than, greater than and low certainty; for the purpose of our analysis, this reading key will be ignored.

```
stormdata$PROPDMGEXP <- as.character(stormdata$PROPDMGEXP)
unique(stormdata$PROPDMGEXP)
```

```
## [1] "K" ""  "M" "B" "0"
```

```
stormdata$PROPDMGEXP[stormdata$PROPDMGEXP %in% c("k", "K")] <- 1000
stormdata$PROPDMGEXP[stormdata$PROPDMGEXP %in% c("m", "M")] <- 1000000
stormdata$PROPDMGEXP[stormdata$PROPDMGEXP %in% c("b", "B")] <- 1000000000
stormdata$PROPDMGEXP[stormdata$PROPDMGEXP %in% c("0", "")] <- 1
stormdata$PROPDMGEXP <- as.numeric(stormdata$PROPDMGEXP)
```

```
stormdata$CROPDMGEXP <- as.character(stormdata$CROPDMGEXP)
unique(stormdata$CROPDMGEXP)
```

```
## [1] "K" ""  "M" "B"
```

```
stormdata$CROPDMGEXP[stormdata$CROPDMGEXP %in% c("k", "K")] <- 1000
stormdata$CROPDMGEXP[stormdata$CROPDMGEXP %in% c("m", "M")] <- 1000000
stormdata$CROPDMGEXP[stormdata$CROPDMGEXP %in% c("b", "B")] <- 1000000000
stormdata$CROPDMGEXP[stormdata$CROPDMGEXP %in% c("0", "")] <- 1
stormdata$CROPDMGEXP <- as.numeric(stormdata$CROPDMGEXP)
```

- Columns "FATALITIES" and "INJURIES" contains the total number of people who suffered from the event; either directly, indirectly or with a certain delay (e.g. fatality occuring a few days after the end of an event). Both are considered as casualties.

```
stormdata <-
  stormdata %>%
   mutate(TOTPROPDMG = PROPDMG * PROPDMGEXP) %>%
      mutate(TOTCROPDMG = CROPDMG * CROPDMGEXP) %>%
```

```
        mutate(CASUALTIES = FATALITIES + INJURIES) %>%
          mutate(DAMAGE = TOTPROPDMG + TOTCROPDMG) %>%
           select(BGN_DATE, EVTYPE, DAMAGE, CASUALTIES) %>%
              filter((DAMAGE+CASUALTIES) != 0)
head(stormdata)
```

```
##     BGN_DATE       EVTYPE DAMAGE CASUALTIES
## 1 1996-01-06 WINTER STORM 418000          0
## 2 1996-01-11      TORNADO 100000          0
## 3 1996-01-11    TSTM WIND   3000          0
## 4 1996-01-11    TSTM WIND   5000          0
## 5 1996-01-11    TSTM WIND   2000          0
## 6 1996-01-18    HIGH WIND 400000          0
```

Data Analysis

In regard of impact on Public Health

In order to tackle the question
Which types of events are most harmful with respect to population health? we will focus only on the casualties per type of event.

```
PHimpact <- filter(stormdata, CASUALTIES != 0)
PHimpact <-
  PHimpact %>%
    group_by(EVTYPE) %>%
      summarize(sum(CASUALTIES))
colnames(PHimpact) <- c("Event", "Sum_of_cas")
PHimpact <- arrange(PHimpact, desc(Sum_of_cas))
```

According to the rule of pareto, we can suppose 20% of the events cause 80% of the casualties, we will thus focus on the top 20% events.

```
pareto1 <- round(length(PHimpact$Sum_of_cas)*0.2)
PHimpact <- PHimpact[1:pareto1,]
PHimpact$Event <- factor(PHimpact$Event, levels = PHimpact$Event[order(PHimpact$Sum_of_cas)])
```

These are the top 28 events.
See "results" for the final graph.

In regard of economic consequences

In order to tackle the question
Which types of events have the greatest economic consequences?
we will focus only on the total damage in dollar per type of event.

```
ECimpact <- filter(stormdata, DAMAGE != 0)
ECimpact <-
  ECimpact %>%
```

```
    group_by(EVTYPE) %>%
      summarize(sum(DAMAGE))
colnames(ECimpact) <- c("Event", "Tot_dam")
ECimpact <- arrange(ECimpact, desc(Tot_dam))
```

Again, we will use the rule of pareto; we will suppose 20% of the events cause 80% of the damage in value, we will thus focus on the top 20% events.

```
pareto2 <- round(length(ECimpact$Tot_dam)*0.2)
ECimpact <- ECimpact[1:pareto1,]
ECimpact$Event <- factor(ECimpact$Event, levels = ECimpact$Event[order(ECimpact$Tot_dam)])
```

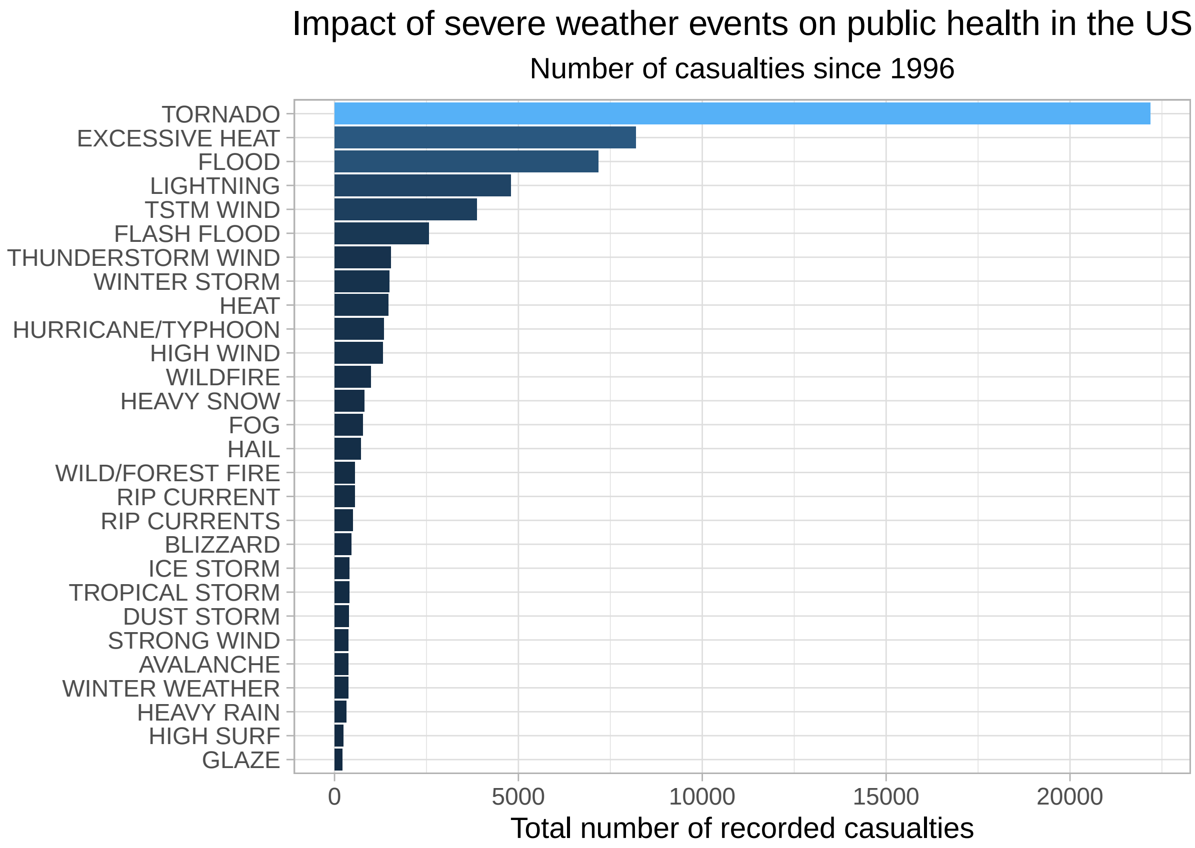
These are the top 37 events.
See "results" for the final graph.

Results

```
library(ggplot2)
```

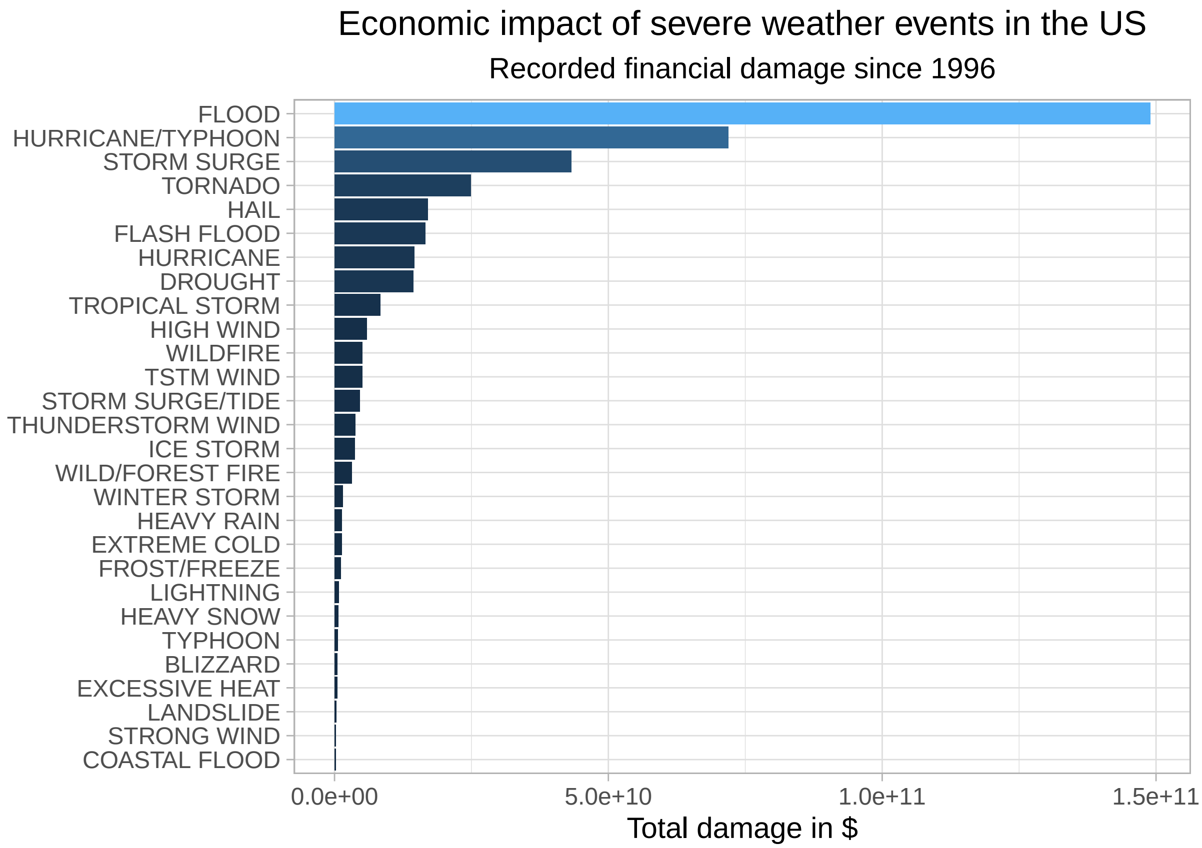
1. Which types of events are most harmful with respect to population health?

```
g <- ggplot(PHimpact)
g + geom_col(aes(Sum_of_cas, Event, fill = Sum_of_cas), show.legend = FALSE) + labs(title = "Impact of s
```



2. Which types of events have the greatest economic consequences?

```
q <- ggplot(ECimpact)
q + geom_col(aes(Tot_dam, Event, fill = Tot_dam), show.legend = FALSE) + labs(title = "Economic impact o
```


Economic impact of severe weather events in the US
Recorded financial damage since 1996
FLOOD
HURRICANE/TYPHOON
STORM SURGE
TORNADO
HAIL
FLASH FLOOD
HURRICANE
DROUGHT
TROPICAL STORM
HIGH WIND
WILDFIRE
TSTM WIND
STORM SURGE/TIDE
THUNDERSTORM WIND
ICE STORM
WILD/FOREST FIRE
WINTER STORM
HEAVY RAIN
EXTREME COLD
FROST/FREEZE
LIGHTNING
HEAVY SNOW
TYPHOON
BLIZZARD
EXCESSIVE HEAT
LANDSLIDE
STRONG WIND
COASTAL FLOOD
0.0e+00
5.0e+10
1.0e+11
1.5e+11
Total damage in $