

Analysis of effects in human health and property damage of major weather events in US

Gabriel General

01-03-2021

SYNOPSIS

Analysis of NOAA Storm Database to study the damage of climatic events.

Many severe weather events can result in human casualties, injuries and property damage and knowing this phenomena can help us to take actions and prevent significant damage to human and property.

Due this, is needed study and analyze the history and data recorded.

This is an analysis of the NOAA Storm Database about severe weather events recorded from 1950 to November 2011.

DATA PROCESSING

1. Getting the data from the National Weather Service

First we set all the parameters needed to get the data file:

```
url = "https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2"
dataDir = "./data"
dataFile = "./data/StormData.csv.bz2"
DataFN = "StormData.csv.bz2"

if (!file.exists(dataDir) || (!file.exists(dataFile))) {
  dir.create(dataDir)
  download.file(url, destfile = "./data/StormData.csv.bz2")
  #unzip(zipfile = DataFileName, exdir = dataDir)
}
```

2. Create a dataset with the data obtained and count the number of NA values.

```
stormData <- read.csv("./data/StormData.csv.bz2")
naValues <- as.numeric(sum(is.na(stormData)))
validValues <- as.numeric(sum(!is.na(stormData)))
perNaValues <- round((naValues*100)/validValues,2)
#perNaValues <- format(round(naValues, 2), nsmall = 2)
```

We observe that exist a **5.52%** of data missed from original dataset.

Columns with invalid data are:

```
sapply(stormData, function(x) sum(is.na(x)))
```

```
##      STATE__ BGN_DATE BGN_TIME TIME_ZONE COUNTY COUNTYNAME STATE
##          0         0         0         0         0         0         0
##      EVTYPE BGN_RANGE BGN_AZI BGN_LOCATI END_DATE END_TIME COUNTY_END
##          0         0         0         0         0         0         0
## COUNTYENDN END_RANGE END_AZI END_LOCATI LENGTH WIDTH F
##    902297         0         0         0         0         0 843563
##      MAG FATALITIES INJURIES PROPDMG PROPDMGEXP CROPDMG CROPDMGEXP
##          0         0         0         0         0         0         0
##      WFO STATEOFFIC ZONENAMES LATITUDE LONGITUDE LATITUDE_E LONGITUDE_
##          0         0         0         47         0         40         0
##      REMARKS REFNUM
##          0         0
```

3. A strategy taken to filter the invalid data was subsetting the original dataset creating two new dataset only with useful data.

3.1 Human casualties: Data in those rows contains at least 1 record with fatalities or injuries called dataDMN

```
dataDMN <- subset(stormData, FATALITIES + INJURIES >0, select = c(STATE,BGN_DATE,COUNTYNAME,STATE,EVTYPE,
numHuman <- round(as.numeric(nrow(dataDMN)))
naDataDMN <- sum(is.na(dataDMN))
```

We've got 2.1929×10^4 records with human casualties and have **NONE** registry with invalid values.

3.2 Property Damage: Data with property damage (in Dollars) called DataDMG

```
dataDMG <- subset(stormData, PROPDMG >0, select = c(STATE,BGN_DATE,COUNTYNAME,STATE,EVTYPE,FATALITIES,I
numPropDMG <- as.numeric(nrow(dataDMG))
naDataDMG <- sum(is.na(dataDMG))
```

We've got 2.39174×10^5 records with economic consequences and have **NONE** registry with invalid values.

Invalid data filtering was not necessary because the casualties and property damage preconditions cleaned up the original dataset.

For example, columns with NA values such as COUNTYENDN, F, or LATITUDE were unnecessary to answer the main questions in this analysis.

This strategy help us to save a lot of time in data pre-processing and memory resources.

RESULTS

Using librarys tidyr and ggplot2 we can identify the most dangerous events.

```
library(dplyr)
library(ggplot2)
```

Most harmful weather events in US

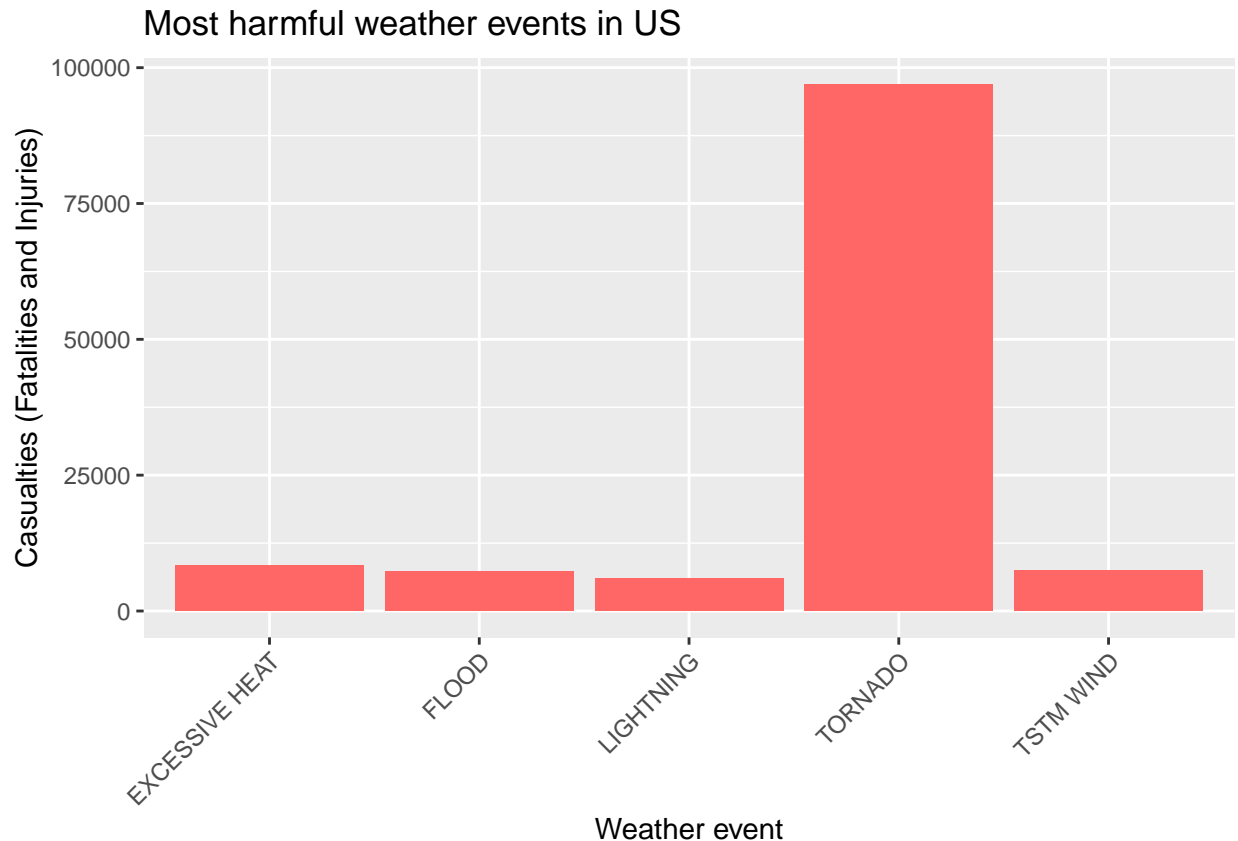
```
human <- dataDMN %>%
  group_by(EVTYPE) %>%
  summarise(CASUALTIES = sum(INJURIES+FATALITIES)) %>%
  top_n(5) %>%
  arrange(desc(CASUALTIES))
```

```
## 'summarise()' ungrouping output (override with '.groups' argument)
```

```
## Selecting by CASUALTIES
```

```
#The top 5 of most harmful event are:
```

```
## # A tibble: 5 x 2
##   EVTYPE      CASUALTIES
##   <chr>         <dbl>
## 1 TORNADO      96979
## 2 EXCESSIVE HEAT  8428
## 3 TSTM WIND     7461
## 4 FLOOD        7259
## 5 LIGHTNING     6046
```



The most harmful weather event in US are **tornadoes** , with more than **90 thousands casualties** between deaths and injuries by far.

Weather events with major economic consequences in US

Property damage is classified by column `PROPDMGEXP` that indicates if value is represented in thousand (K) or million of dollars (M). So, to summarize this we must grouping by value expression an then sum by total.

```
miles <- dataDMG %>%
  filter(PROPDMGEXP == "K") %>%
  group_by(EVTYPE) %>%
  summarise(DAMAGE = sum(PROPDMG)/1000)
```

'summarise()' ungrouping output (override with '.groups' argument)

```
millones <- dataDMG %>%
  filter(PROPDMGEXP == "M") %>%
  group_by(EVTYPE) %>%
  summarise(DAMAGE = sum(PROPDMG))
```

'summarise()' ungrouping output (override with '.groups' argument)

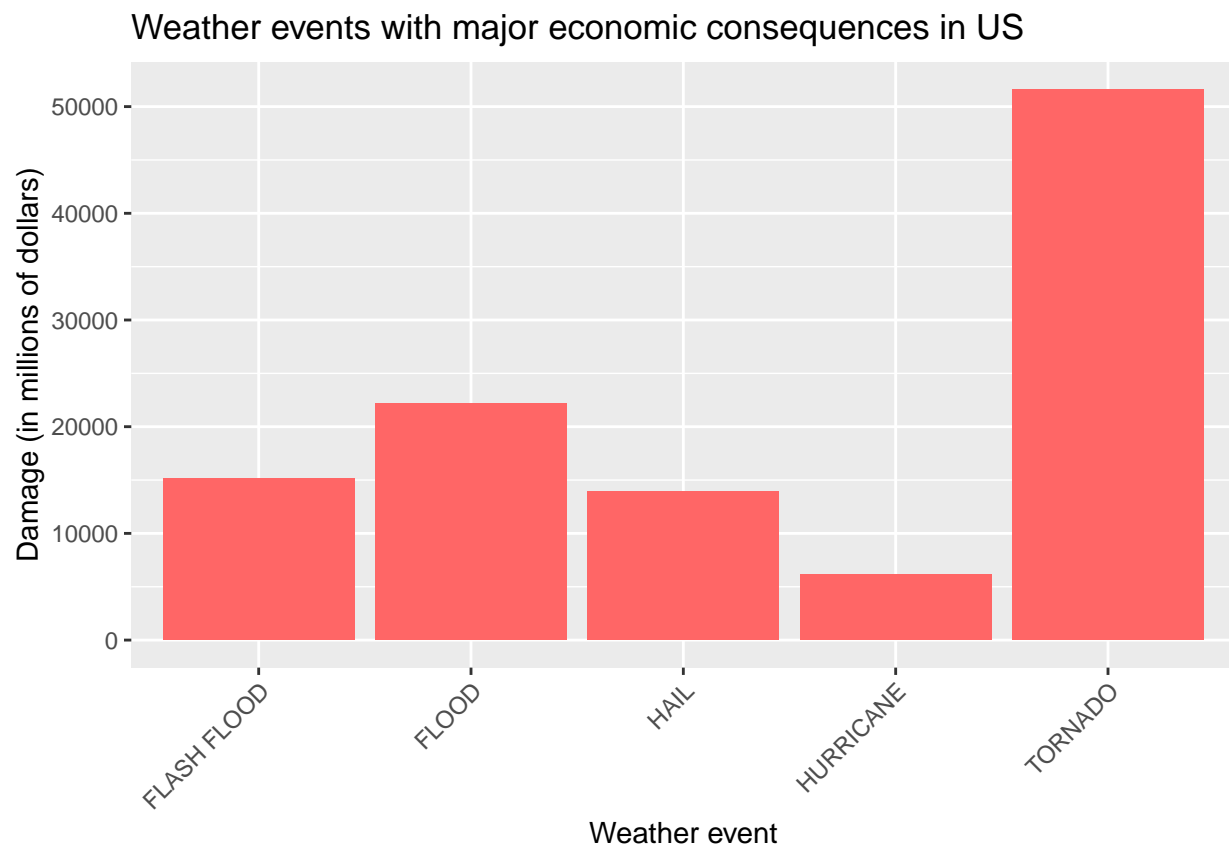
```
econ <- bind_rows(miles, millones) %>%
  group_by(EVTYPE) %>%
  summarise_all(funs(sum(., na.rm = TRUE))) %>%
  top_n(5)
```

```
## Warning: 'funs()' is deprecated as of dplyr 0.8.0.
## Please use a list of either functions or lambdas:
##
##   # Simple named list:
##   list(mean = mean, median = median)
##
##   # Auto named with 'tibble::lst()':
##   tibble::lst(mean, median)
##
##   # Using lambdas
##   list(~ mean(., trim = .2), ~ median(., na.rm = TRUE))
## This warning is displayed once every 8 hours.
## Call 'lifecycle::last_warnings()' to see where this warning was generated.
```

```
## Selecting by DAMAGE
```

```
#The top 5 of event with major economic consequences in US:
```

```
## # A tibble: 5 x 2
##   EVTYPE      DAMAGE
##   <chr>      <dbl>
## 1 TORNADO    51626.
## 2 FLOOD     22158.
## 3 FLASH FLOOD 15141.
## 4 HAIL      13927.
## 5 HURRICANE   6168.
```



We can see that the climate event with the greatest economic consequences are **tornadoes**, with more than **50 billion dollars** in damages, followed by floods, hail and hurricanes. A strategy to prevent more damage in the states most prone to this climate phenomenon, would be to develop forecasting systems and improve communications and coordinated measures to avoid more human and economic damage.