

Analyzing the NOAA storm database

Geraldo Barbosa do Amarante

5 de dezembro de 2017

Synopsis

Damage caused by storms, in localities and cities, causes numerous catastrophes. The effects of these disasters, as well as the emotional and affective, are the economic ones. The families directly involved lose house, car, belongings and, at least, they need to spend some money saved to rebuild some of what has been damaged. In addition to leaving a trail of destruction, with trees felled, homes and buildings destroyed and floods, storms can bring serious consequences for human health. Therefore, we must remain alert for all the problems that this phenomenon of nature can cause. The damage can be huge if we are not prepared. This project involves the analysis of the U.S. NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION's (NOAA) database. This database information on storms and events caused by climate change in the US, including when they occurred, where and estimates of fatalities and injuries caused by these events.

In this analysis, we intend to answer basically two questions:

- 1 - Across the United States, which types of events (as indicated in the EVTYPE variable) are most harmful with respect to population health?
- 2 - Across the United States, which types of events have the greatest economic consequences?

Data Processing

```
##-----  
## Load Libraries ----  
##-----  
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 3.4.3
```

```
##  
## 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
```

```
library(lubridate)
```

```
## Warning: package 'lubridate' was built under R version 3.4.3
```

```
##  
## Attaching package: 'lubridate'
```

```
## The following object is masked from 'package:base':  
##  
##   date
```

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.4.3
```

```
library(sqldf)
```

```
## Warning: package 'sqldf' was built under R version 3.4.3
```

```
## Loading required package: gsubfn
```

```
## Warning: package 'gsubfn' was built under R version 3.4.3
```

```
## Loading required package: proto
```

```
## Warning: package 'proto' was built under R version 3.4.3
```

```
## Loading required package: RSQLite
```

```
library(knitr)
```

```
## Warning: package 'knitr' was built under R version 3.4.3
```

```
##-----  
## Defining file path -----  
##-----  
setwd('C:/Amarante/ReproducibleResearch/ArquivoDeDados')  
## Reading file  
arquivo <- read.csv("./StormData.csv", sep=";", header=T)
```

```
## - Sql query for first question  
comando_sql <- "select EVTYPE as Event,sum(FATALITIES) as Fatalites,sum(INJURIES) as lesoes from  
m arquivo  
      group by EVTYPE order by sum(FATALITIES) desc,sum(INJURIES) desc limit 10"  
# Execução do comando  
resultado_sql01 <-sqldf::sqldf(comando_sql)
```

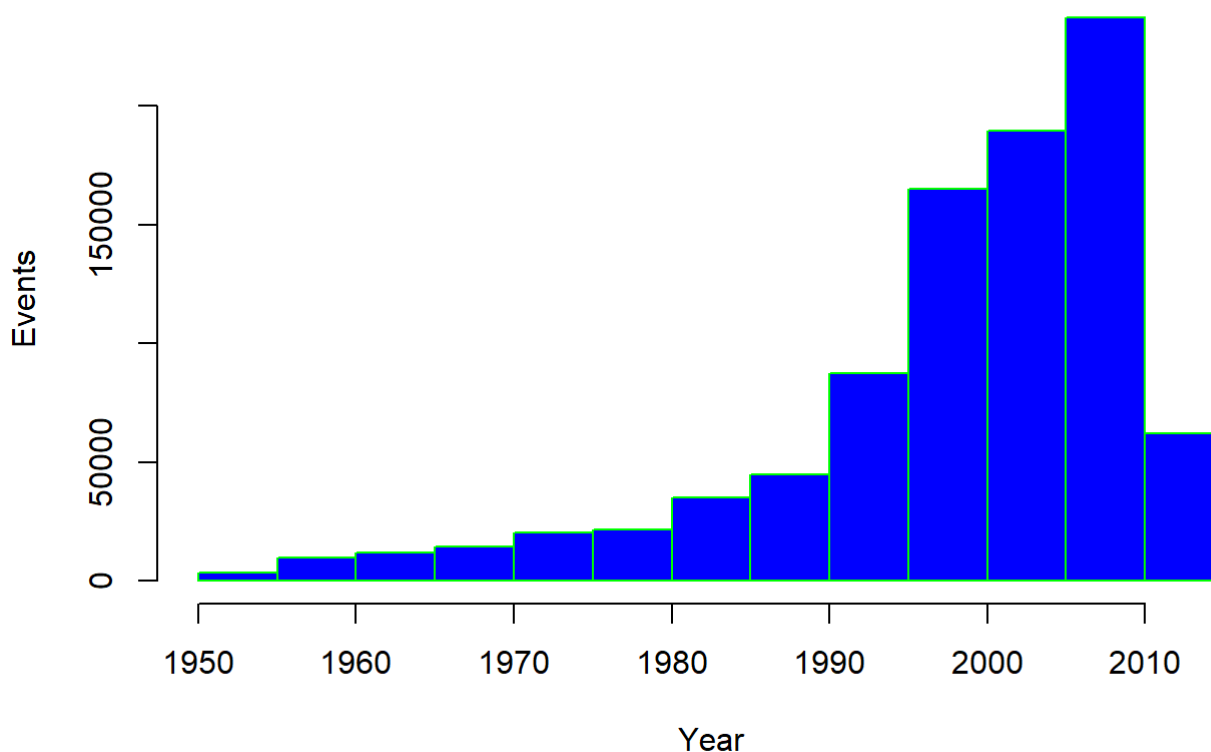
```
comando_sql <- "select EVTYPE as Event,sum(PROPDMG)/1000 as Property_Damage from arquivo group  
by EVTYPE order by sum(PROPDMG) desc limit 10"  
# Execução do comando  
resultado_sql02 <-sqldf(comando_sql)
```

Results

This first graphics show the ocorrency of storms and weather events in the United States by year, from 1950 to 2010.

```
hist(year(as.Date(arquivo$BGN_DATE, '%m/%d/%Y')), col = "blue", main="Major storms and weather events in the United States",xlab = "Year",ylab="Events",border="green")
```

Major storms and weather events in the United States



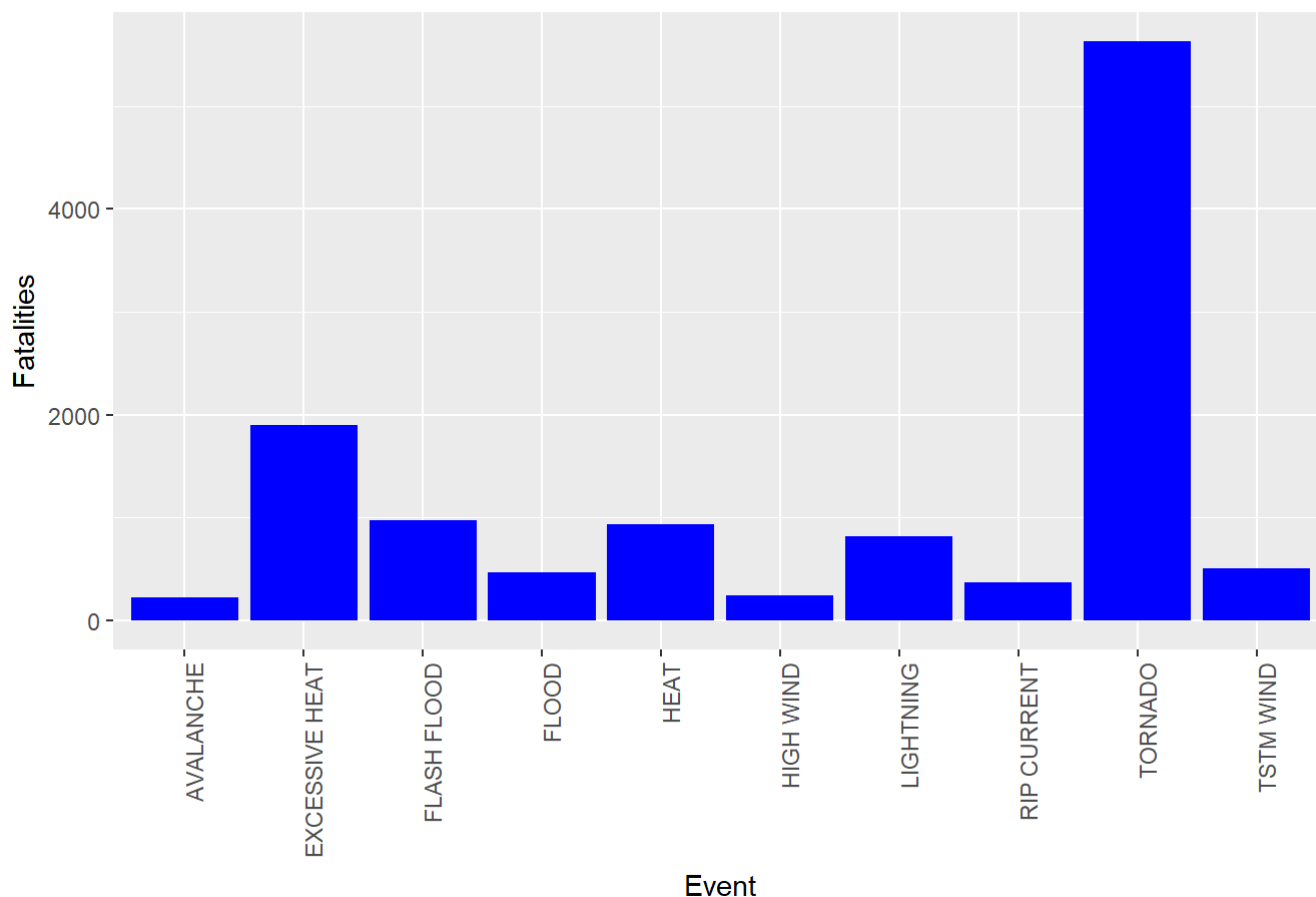
```
resultado_sql02 <- sqldf(comando_sql)
```

1 - Across the United States, which types of events (as indicated in the EVTYPE variable) are most harmful with respect to population health?

In this analysis, it is observed that the tornado event is the most responsible for the occurrence of fatalities, reaching twice the number of occurrences of EXCESSIVE HEAT, which is the second highest occurrence among the 10 largest.

```
ggplot(resultado_sql01, aes(x = Event, y = Fatalites)) + geom_bar(stat = "identity", fill = "blue") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1)) +
  xlab("Event") + ylab("Fatalities") + ggtitle("The greater 10 Ocorrency of fatalities by Events")
```

The greater 10 Ocorrecny of fatalities by Events



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

Analysing this second graphics, it is observed that the tornado event is the most responsible for the occurrence of property damage, reaching twice the number of occurrences of FLASHFLOOD, the second highest in the occurrence of 10 more frequent.

```
ggplot(resultado_sql02, aes(x = Event, y = Property_Damage)) + geom_bar(stat = "identity", fill = "green") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1)) +
  xlab("Event") + ylab("Property Damage X 1000") + ggtitle("The greater 10 Ocorrecny of Property Damage by Events")
```

The greater 10 Ocorrency of Property Damage by Events

