

---

---

title: Analysis of health and property damage for Weather Events Using NOAA Storm Database

author: "Heloísa Garcia Claro"

date: "Sunday, March 13, 2016"

output: html\_document

---

##Synopsis##

##Analysis using The NOAA Storm Database to evaluate weather events in terms of damage for health and property. Goal is to answer the questions:##

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

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

#Assignment 2#

###Install package knitr###

install.packages("knitr")

###Load Knitr###

library(knitr)

###set working directory###

setwd("C:/Users/Vaio/Desktop")

####Read data####

dataweather <- read.csv("C:/Users/Vaio/Desktop/repdata-data-StormData.csv.bz2")

####Check data####

```

View(dataweather)

##Data Processing##

####Needed data transformations####

####verify number of unique event types####

length(unique(dataweather$EVTYPE))

####change letters types to tolwer####

EVTYPE_transf <- tolower(dataweather$EVTYPE)

####remove spaces, dots and other characters other than letters and numbers####

EVTYPE_transf <- gsub("[[:blank:][:punct:]]+", " ", EVTYPE_transf)

length(unique(EVTYPE_transf))

####replace tstm wind to thunderstorm wind####

dataweather$EVTYPE[dataweather$EVTYPE=="tstm wind"] <- "thunderstorm wind"

##Now the data is transformed to reduce and recategorize the same events that were
reported separately##

####change original data frame####

dataweather$EVTYPE <- EVTYPE_transf

####Get the data dictionary####

str(dataweather)

####Load needed packages to explore the data####

####Lubridate to deal with dates and times####

library(lubridate)

####dplyr for easy data manipulation####

library(dplyr)

```

```
##Transform data##
```

```
####Create a variable to separate data by year####
```

```
dataweather$year <- year(as.Date(dataweather$BGN_DATE, '%m/%d/%Y'))
```

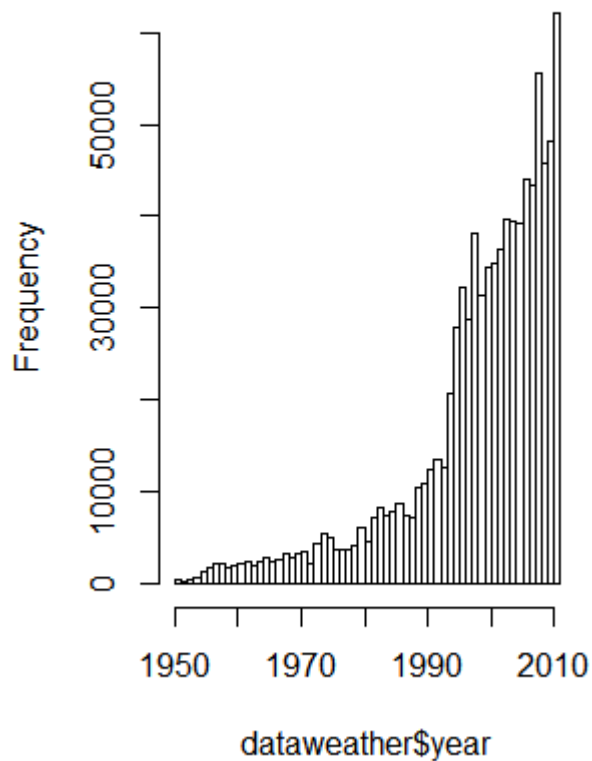
```
##Results##
```

```
##Explore data##
```

```
####exploring the data using histogram plot for time series of data from storms####
```

```
hist(dataweather$year, 62)
```

## Histogram of dataweather\$year



#as it can be seen, the data corresponds to the affirmation from the assignment that "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"#

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

```
####Load package sqldf to use sql language on Rstudio.####
```

```
library(sqldf)
```

```
sql <- "
```

```
select
```

```
  EVTYPE as type_event,
```

```
  sum(FATALITIES) as cas_fatal,
```

```
  sum(INJURIES) as cas_injuries
```

```
from
```

```
  dataweather
```

```
where
```

```
  year >= 2000
```

```
group by
```

```
  EVTYPE
```

```
order by
```

```
  sum(FATALITIES) desc,
```

```
  sum(INJURIES) desc
```

```
"
```

```
tbl_ds <-sqldf::sqldf(sql)
```

```
####Load package tcltk####
```

```
library(tcltk)
```

```
#####Explore table for cas_fatal and cas_injuries as types of event####
```

```
library(knitr)
```

```
knitr::kable(tbl_ds[1:10,])
```

type_event	cas_fatal	cas_injuries
tornado	1193	15213
excessive heat	1013	3708

flash flood	600	812
lightning	466	2993
rip current	340	208
flood	266	315
thunderstorm wind	246	3153
heat	231	1222
avalanche	179	126
high wind	131	677

#As it can be seen on the table, the ten most harmful events from 2000 on are tornado, excessive heat, flash flood, lightning, RIP current, flood, thunderstorm wind, heat, avalanche and high wind with respect to population health.#

#Question 2: Across the United States, which types of events have the greatest economic consequences?#

```
library(sqldf)

sql <- "
  select
    EVTYPE as type_event,
    sum(PROPDMG)
  from
    dataweather
  where
    YEAR >= 2000
  group by
    EVTYPE
  order by
    sum(PROPDMG) desc
"

tbl_ds <- sqldf(sql)

####Explore data for property_damage by type_event####

names(tbl_ds) <- c('type_event', 'property_damage')

library(knitr)

knitr::kable(tbl_ds[1:10,])
```

type_event	property_damage
thunderstorm wind	1673785.58
flash flood	999333.42
tornado	907111.70
flood	671747.56
hail	452533.47
lightning	395884.69
high wind	247108.53
winter storm	97093.93
wildfire	83007.34
heavy snow	64675.48

#As it can be seen on the table, the types of event that mostly brought damage to properties were thunderstorm wind, flashflood, tornado, flood, hail, lightning, highwind, winter storm, wildfire and heavy snow.#

#Conclusions: , the ten most harmful events from 2000 on are tornado, excessive heat, flash flood, lightning, RIP current, flood, thunderstorm wind, heat, avalanche and high wind with respect to population health and the types of event that mostly brought damage to properties were thunderstorm wind, flashflood, tornado, flood, hail, lightning, highwind, winter storm, wildfire and heavy snow#