Impacts on the economy and public health by severe weather events in the US

Paulo Cardoso

December 23, 2015

Set up

```
options(scipen = 1)  # remove scientific notations
```

Synopsis

This work aims to carry out an analysis of the impacts of extreme climatic events on the economy and public health in the United States. The data used in this analysis are data collected by the US National Oceanic and Atmospheric Administration's (NOAA) and cover the 1950 period - 2011. In this the variables data set taken into consideration in this analysis can be cited: fatalities, injuries and property damage (in US dollars).The data used is available here (https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2 (https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2)) and the documenation if available here (https://d396qusza40orc.cloudfront.net/repdata%2Fpeer2_doc%2Fpd01016005curr.pdf (https://d396qusza40orc.cloudfront.net/repdata%2Fpeer2_doc%2Fpd01016005curr.pdf))

Data Processing

This section consists of the loading of data and its preprocessing, going through a quick exploratorioa analysis of data in order to conceptualize the dataset, thus providing the presentation of the variables utilized in the study.

Data Load

After downloading the data by the link provided above, the data is loaded into the variable Datastorm.

```
dataStorm <- read.csv(bzfile("repdata_data_StormData.csv.bz2"))
```

Exploratory Data Analisys

As can be seen below the dataset has 902,297 rows and 37 columns, and is also available a sample of the dataset.

```
dim(dataStorm)
```

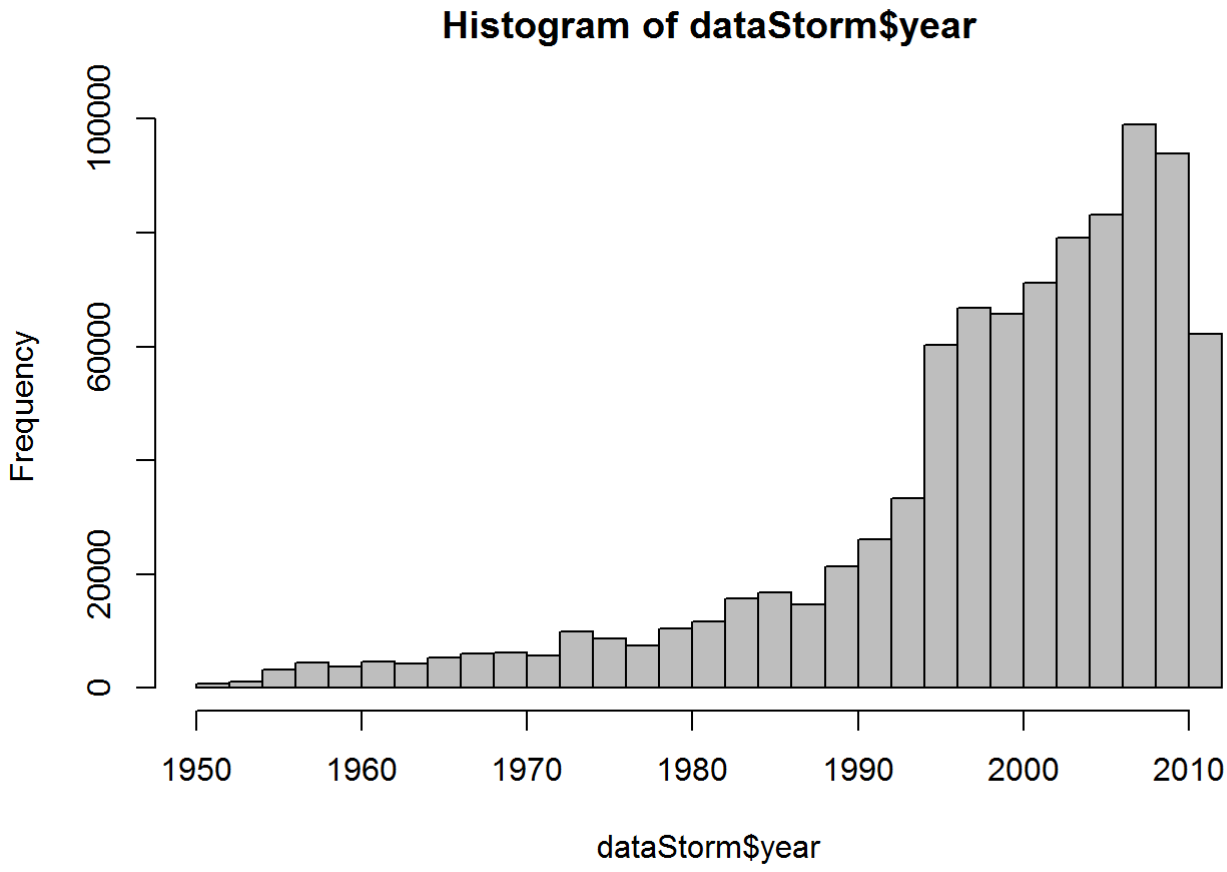
```
## [1] 902297     37
```

```
head(dataStorm, n = 3)
```

```
##   STATE__           BGN_DATE BGN_TIME TIME_ZONE COUNTY COUNTYNAME STATE
## 1       1 4/18/1950 0:00:00     0130       CST     97     MOBILE    AL
## 2       1 4/18/1950 0:00:00     0145       CST      3    BALDWIN    AL
## 3       1 2/20/1951 0:00:00     1600       CST     57    FAYETTE    AL
##    EVTYPE BGN_RANGE BGN_AZI BGN_LOCATI END_DATE END_TIME COUNTY_END
## 1 TORNADO         0                                               0
## 2 TORNADO         0                                               0
## 3 TORNADO         0                                               0
##   COUNTYENDN END_RANGE END_AZI END_LOCATI LENGTH WIDTH F MAG FATALITIES
## 1         NA         0                      14.0   100 3   0          0
## 2         NA         0                       2.0   150 2   0          0
## 3         NA         0                       0.1   123 2   0          0
##   INJURIES PROPDMG PROPDMGEXP CROPDMG CROPDMGEXP WFO STATEOFFIC ZONENAMES
## 1       15    25.0          K       0
## 2        0     2.5          K       0
## 3        2    25.0          K       0
##   LATITUDE LONGITUDE LATITUDE_E LONGITUDE_ REMARKS REFNUM
## 1     3040      8812       3051       8806              1
## 2     3042      8755          0          0              2
## 3     3340      8742          0          0              3
```

As can be noted in the histogram, the frequency distribution of the data inputs have a higher intensity of the late ’90s to now.

```
if (dim(dataStorm)[2] == 37) {dataStorm$year <- as.numeric(format(as.Date(dataStorm$B
GN_DATE, format = "%m/%d/%Y %H:%M:%S"), "%Y")) }
hist(dataStorm$year, breaks = 30, col = "gray")
```



Normalizing names of events.

```
x <- dataStorm[,c("EVTYPE", "FATALITIES", "INJURIES", "PROPDMG")]

x$EVTYPE <- gsub("^HEAT$", "EXCESSIVE HEAT", x$EVTYPE)
x$EVTYPE <- gsub("^TSTM WIND$", "THUNDERSTORM WIND", x$EVTYPE)
x$EVTYPE <- gsub("^THUNDERSTORM WIND$", "THUNDERSTORM WINDS", x$EVTYPE)
```

Aggregating the top 10 fatalities causes.

```
fatal<- aggregate(x$FATALITIES, by=list(x$EVTYPE), FUN=sum, na.rm=TRUE)
colnames(fatal) = c("event.type", "fatality.total")
f <- fatal[order(-fatal$fatality.total),]
t10fatal <- f[1:10,]
t10fatal$event.type <- factor(t10fatal$event.type, levels = t10fatal$event.type, orde
red=TRUE)
t10fatal
```

```
##             event.type fatality.total
## 832            TORNADO           5633
## 130     EXCESSIVE HEAT           2840
## 153        FLASH FLOOD            978
## 463          LIGHTNING            816
## 784 THUNDERSTORM WINDS            701
## 170              FLOOD            470
## 584        RIP CURRENT            368
## 358          HIGH WIND            248
## 19           AVALANCHE            224
## 969       WINTER STORM            206
```

Aggregating the top 10 injuries causes.

```
injur<- aggregate(x$INJURIES, by=list(x$EVTYPE), FUN=sum, na.rm=TRUE)
colnames(injur) = c("event.type", "injury.total")
injuries.sorted <- injur[order(-injur$injury.total),]
t10inj <- injuries.sorted[1:10,]
t10inj$event.type <- factor(t10inj$event.type, levels=t10inj$event.type, ordered=TRUE
)
t10inj
```

```
##             event.type injury.total
## 832            TORNADO        91346
## 784 THUNDERSTORM WINDS         9353
## 130     EXCESSIVE HEAT         8625
## 170              FLOOD         6789
## 463          LIGHTNING         5230
## 426          ICE STORM         1975
## 153        FLASH FLOOD         1777
## 244               HAIL         1361
## 969       WINTER STORM         1321
## 410  HURRICANE/TYPHOON         1275
```

Aggregating the top 10 property damage causes.

```
propdmg <- aggregate( x$PROPDMG, by=list(x$EVTYPE), FUN=sum, na.rm=TRUE)
colnames(propdmg) = c("event.type", "prop.dmg.total")
prop.dmg.sorted <- propdmg[order(-propdmg$prop.dmg.total),]
t10pd <- prop.dmg.sorted[1:10,]
t10pd$event.type <- factor(t10pd$event.type, levels=t10pd$event.type, ordered=TRUE)
t10pd
```

```
##             event.type prop.dmg.total
## 832            TORNADO     3212258.16
## 784 THUNDERSTORM WINDS     2659102.96
## 153        FLASH FLOOD     1420124.59
## 170              FLOOD      899938.48
## 244               HAIL      688693.38
## 463          LIGHTNING      603351.78
## 358          HIGH WIND      324731.56
## 969       WINTER STORM      132720.59
## 309         HEAVY SNOW      122251.99
## 954           WILDFIRE       84459.34
```

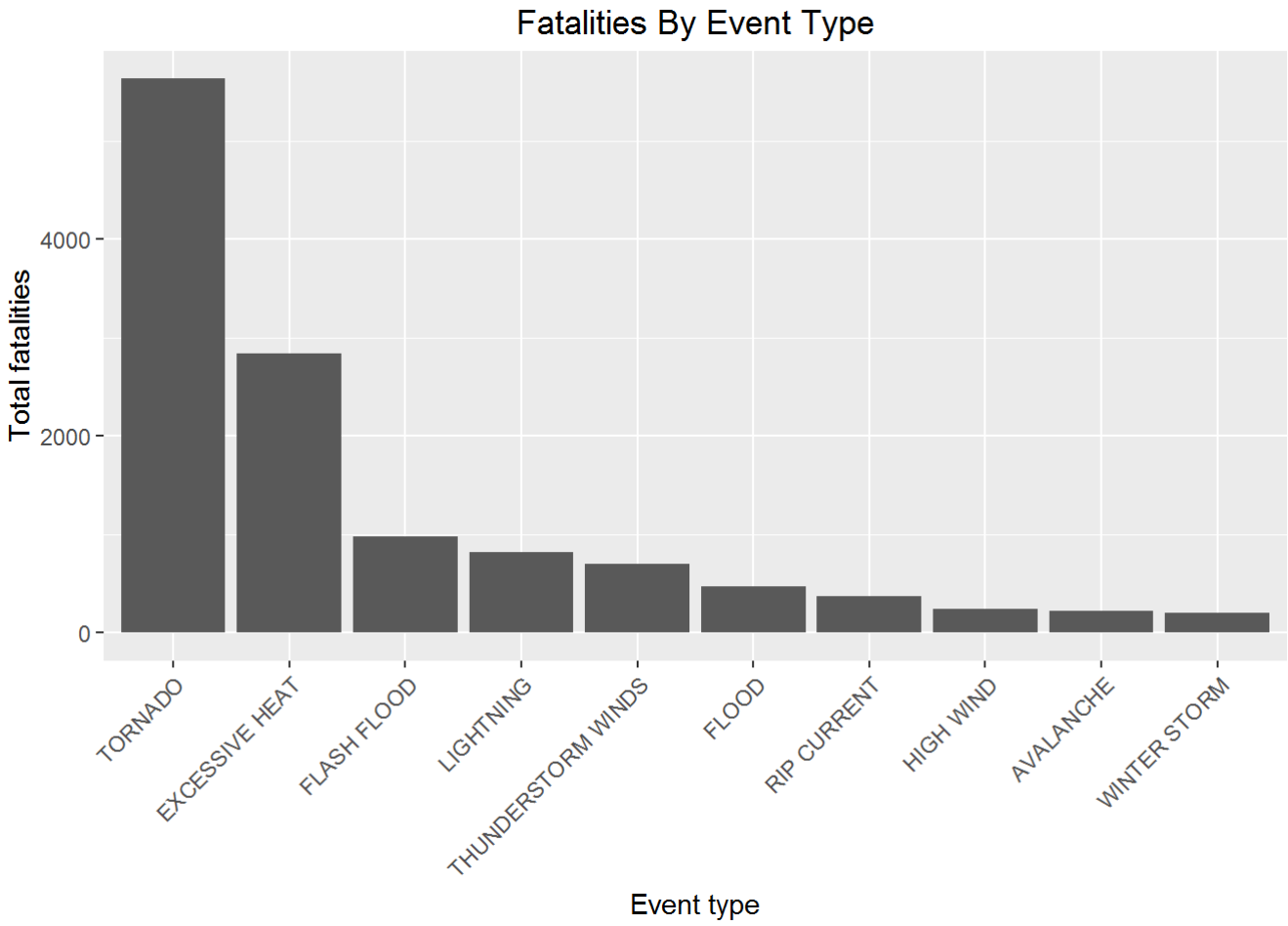
Results

The damages in the area of public health are the following:

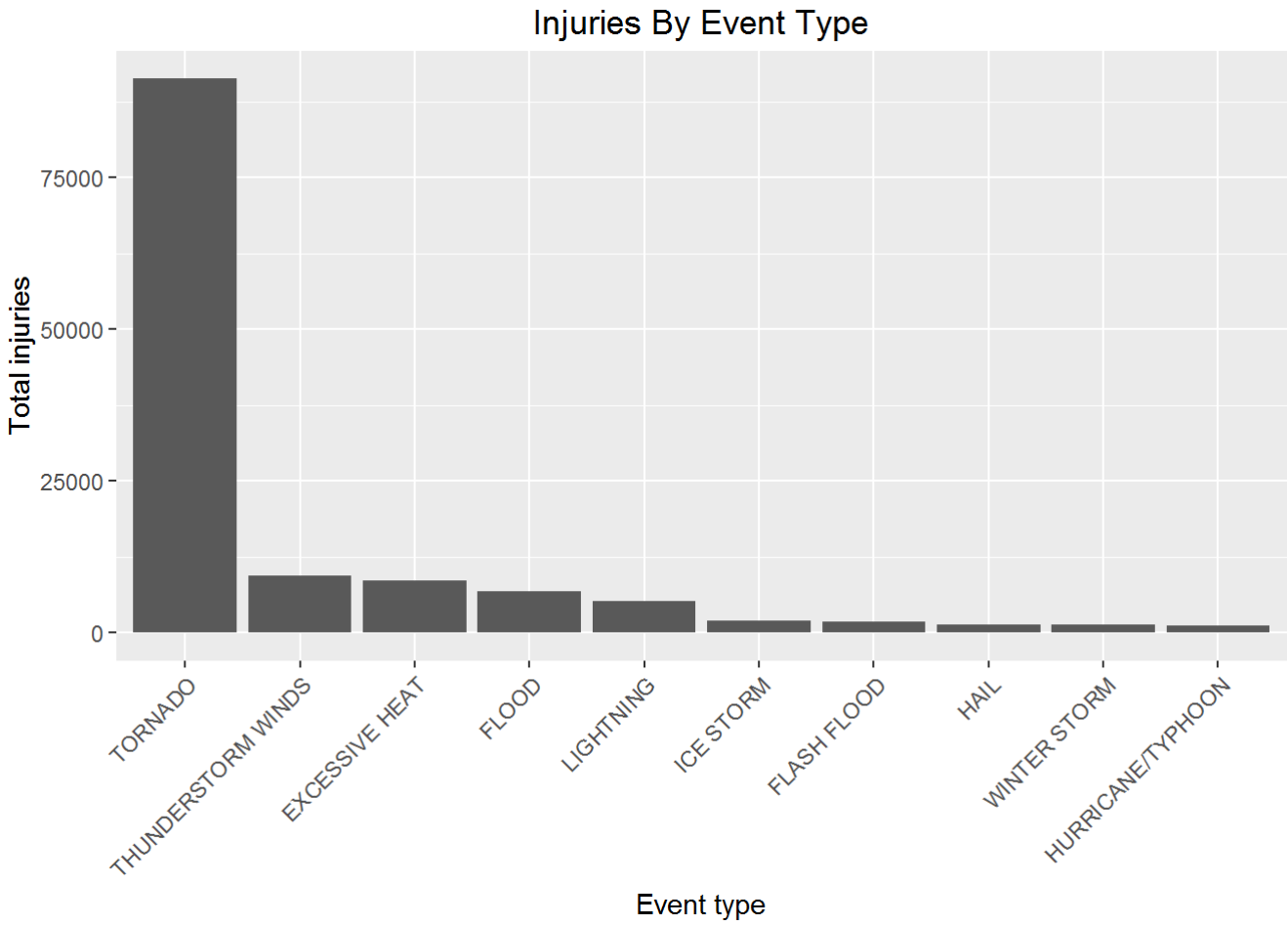
```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.2.3
```

```
ggplot(data=t10fatal, aes(x=event.type, y=fatality.total)) +
    geom_bar(stat="identity") + xlab("Event type") + ylab("Total fatalities") +
    ggtitle("Fatalities By Event Type") +
    theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

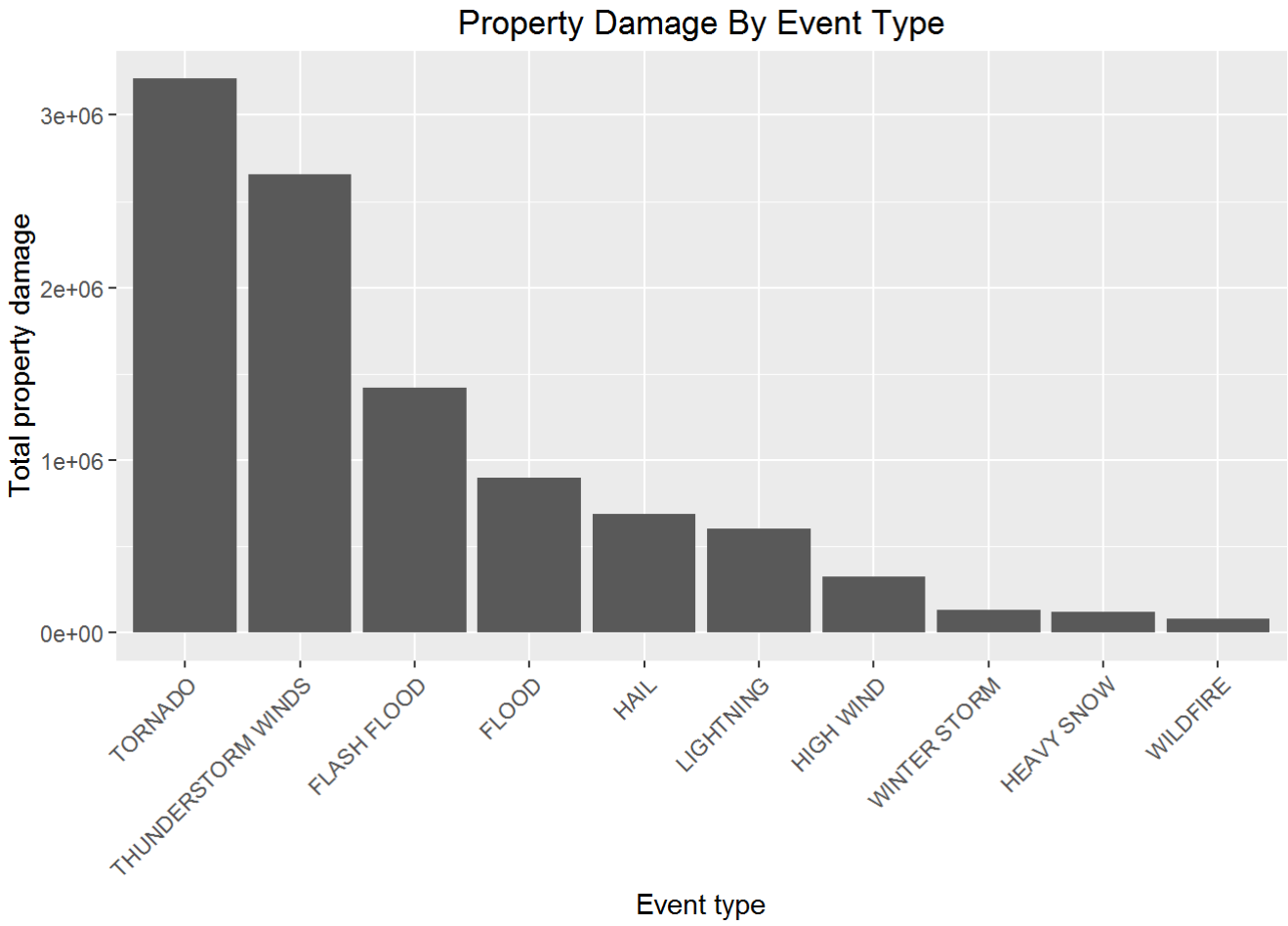



```
ggplot(data=t10inj, aes(x=event.type, y=injury.total)) +
    geom_bar(stat="identity") + xlab("Event type") + ylab("Total injuries") +
    ggtitle("Injuries By Event Type") +
    theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



Finally the economic damage are:

```
ggplot(data=t10pd, aes(x=event.type, y=prop.dmg.total)) +
    geom_bar(stat="identity") + xlab("Event type") + ylab("Total property damage") +
ggtitle("Property Damage By Event Type") +
    theme(axis.text.x = element_text(angle = 45, hjust = 1))
```


Property Damage By Event Type
Total property damage
3e+06
2e+06
1e+06
0e+00
TORNADO
THUNDERSTORM WINDS
FLASH FLOOD
FLOOD
HAIL
LIGHTNING
HIGH WIND
WINTER STORM
HEAVY SNOW
WILDFIRE
Event type