Peer Assesment2: Natural Dissasters

Sotero

Saturday, November 15, 2014

Data Processing

First we download the data.

```
download.file("http://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2","C:/Users/Soter

data <-read.csv("data.csv", header= T )
```

In this part we find the total number of fatalities for each type of disaster.

```
str(data)
```

```
## 'data.frame':    902297 obs. of  37 variables:
##  $ STATE__   : num  1 1 1 1 1 1 1 1 1 1 ...
##  $ BGN_DATE  : Factor w/ 16335 levels "1/1/1966 0:00:00",..: 6523 6523 4242 11116 2224 2224 2260 383
##  $ BGN_TIME  : Factor w/ 3608 levels "00:00:00 AM",..: 272 287 2705 1683 2584 3186 242 1683 3186 318
##  $ TIME_ZONE : Factor w/ 22 levels "ADT","AKS","AST",..: 7 7 7 7 7 7 7 7 7 7 ...
##  $ COUNTY    : num  97 3 57 89 43 77 9 123 125 57 ...
##  $ COUNTYNAME: Factor w/ 29601 levels "","5NM E OF MACKINAC BRIDGE TO PRESQUE ISLE LT MI",..: 13513
##  $ STATE     : Factor w/ 72 levels "AK","AL","AM",..: 2 2 2 2 2 2 2 2 2 2 ...
##  $ EVTYPE    : Factor w/ 985 levels "   HIGH SURF ADVISORY",..: 834 834 834 834 834 834 834 834 834 8
##  $ BGN_RANGE : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ BGN_AZI   : Factor w/ 35 levels "","  N"," NW",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ BGN_LOCATI: Factor w/ 54429 levels "","- 1 N Albion",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ END_DATE  : Factor w/ 6663 levels "","1/1/1993 0:00:00",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ END_TIME  : Factor w/ 3647 levels "","  0900CST",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ COUNTY_END: num  0 0 0 0 0 0 0 0 0 0 ...
##  $ COUNTYENDN: logi  NA NA NA NA NA NA ...
##  $ END_RANGE : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ END_AZI   : Factor w/ 24 levels "","E","ENE","ESE",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ END_LOCATI: Factor w/ 34506 levels "","- .5 NNW",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ LENGTH    : num  14 2 0.1 0 0 1.5 1.5 0 3.3 2.3 ...
##  $ WIDTH     : num  100 150 123 100 150 177 33 33 100 100 ...
##  $ F         : int  3 2 2 2 2 2 2 1 3 3 ...
##  $ MAG       : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ FATALITIES: num  0 0 0 0 0 0 0 0 1 0 ...
##  $ INJURIES  : num  15 0 2 2 2 6 1 0 14 0 ...
##  $ PROPDMG   : num  25 2.5 25 2.5 2.5 2.5 2.5 2.5 25 25 ...
##  $ PROPDMGEXP: Factor w/ 19 levels "","-","?","+",..: 17 17 17 17 17 17 17 17 17 17 ...
##  $ CROPDMG   : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ CROPDMGEXP: Factor w/ 9 levels "","?","0","2",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ WFO       : Factor w/ 542 levels "","  CI","$AC",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ STATEOFFIC: Factor w/ 250 levels "","ALABAMA, Central",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ ZONENAMES : Factor w/ 25112 levels "","
##  $ LATITUDE  : num  3040 3042 3340 3458 3412 ...
```

```
##  $ LONGITUDE : num  8812 8755 8742 8626 8642 ...
##  $ LATITUDE_E: num  3051 0 0 0 0 ...
##  $ LONGITUDE_: num  8806 0 0 0 0 ...
##  $ REMARKS   : Factor w/ 436781 levels "","-2 at Deer Park\n",..: 1 1 1 1 1 1 1 1 1 ...
##  $ REFNUM    : num  1 2 3 4 5 6 7 8 9 10 ...
```

```
head(data)
```

```
##   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
## 4       1   6/8/1951 0:00:00     0900       CST     89    MADISON    AL
## 5       1 11/15/1951 0:00:00     1500       CST     43    CULLMAN    AL
## 6       1 11/15/1951 0:00:00     2000       CST     77 LAUDERDALE    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
## 4 TORNADO         0                                               0
## 5 TORNADO         0                                               0
## 6 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
## 4         NA         0                     0.0   100 2   0          0
## 5         NA         0                     0.0   150 2   0          0
## 6         NA         0                     1.5   177 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
## 4        2     2.5          K       0
## 5        2     2.5          K       0
## 6        6     2.5          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
## 4     3458      8626          0          0              4
## 5     3412      8642          0          0              5
## 6     3450      8748          0          0              6
```

```
fatal  <- tapply(data$FATALITIES , data$EVTYPE , sum, na.rm= T)


fataldf <- data.frame(EVTYPE = names(fatal) , FATALITIES = fatal     )
```

No we find the disaster type with the most number of fatalities.

```
index <- which.max(fataldf$FATALITIES)

fataldf$EVTYPE[index]
```

```
## [1] TORNADO
## 985 Levels:    HIGH SURF ADVISORY  COASTAL FLOOD ... WND
```

Tornadoes killed a total of

```
max(fataldf$FATALITIES)
```

```
## [1] 5633
```

people.

Now we will look for the disaster type that causes the most injuries

```
injur <- tapply(data$INJURIES,data$EVTYPE,sum, na.rm = T  )

injurdf <- data.frame(EVTYPE = names(injur),INJURIES = injur   )

indexi <- which.max(injurdf$INJURIES)

injurdf$EVTYPE[indexi]
```

```
## [1] TORNADO
## 985 Levels:    HIGH SURF ADVISORY  COASTAL FLOOD ... WND
```

Tornadoes have caused

```
max(injurdf$INJURIES)
```

```
## [1] 91346
```

injuries

Plots

```
library("ggplot2", lib.loc="~/R/win-library/3.1")




data$BGN_DATE <-strptime(data$BGN_DATE, format = "%m/%d/%Y %H:%M:%S" )

tordata <- subset(data,EVTYPE == "TORNADO"  )



# here we compute the total tornado fatalities per bgn_date. To do this we extrat the year from the #bg
```

```
tordata$BGN_DATE <- format(tordata$BGN_DATE,"%Y")



toryear <- tapply(tordata$FATALITIES,tordata$BGN_DATE,sum)

toryeardf <- data.frame(BGN_DATE = names(toryear),FATALITIES = toryear   )


toryeardf$BGN_DATE <- as.character(toryeardf$BGN_DATE )

toryeardf$BGN_DATE <-as.numeric(toryeardf$BGN_DATE )


#plot(toryeardf$BGN_DATE , toryeardf$FATALITIES,xlab="Year",ylab="Fatalities",type="l",main=  "Tornados

#dev.copy(png,file="tornadofatal.png")

#dev.off()
```

Now examine any financial damage

‘‘‘

In this part we find the total number of property damages for each type of disaster.

```
dame  <- tapply(data$PROPDMG , data$EVTYPE , sum, na.rm= T)


damagesdf <- data.frame(EVTYPE = names(dame) , PROPDMG = dame     )
```

Now we find the disaster type with the most property damages

```
indexd <- which.max(damagesdf$PROPDMG)

damagesdf$EVTYPE[indexd]
```

```
## [1] TORNADO
## 985 Levels:    HIGH SURF ADVISORY  COASTAL FLOOD ... WND
```

Tornados have damaged

```
max(damagesdf$PROPDMG)
```

```
## [1] 3212258
```

properties.

```r
# here we find the total number of properties damaged each year by tornados

dmy <- tapply(tordata$PROPDMG,tordata$BGN_DATE,sum)

damprop <- data.frame(BGN_DATE = names(dmy),PROPDMG = dmy    )


damprop$BGN_DATE <- as.character(damprop$BGN_DATE )

damprop$BGN_DATE <-as.numeric(damprop$BGN_DATE )


#plot(damprop$BGN_DATE,damprop$PROPDMG  )

#with(data=damprop,plot(BGN_DATE,PROPDMG ) )
```