

Relationship of severe weather events on health and economic - 1950/2010

Oscar Sanchez

22 de octubre de 2015

ANALYSIS

Sinopsis

In order to give accurate information to prepare them for severe weather events and like fundamental base the prioritization of resources for different types of events. For getting the 10 first, I only used the information of the last 11 years to know the events most dangerous for this time period, but, after for the plots I use all data.

In the first question, I ordered the events with more injuries and fatalities. In the second question, I ordered the events with more crop damage and property damages, after, I choose the 3 first events and I drew them from 1950 to 2010, also, I united the years in period of three years for example 1950, 1951, 1952, 1953 are 1950,1950,1950,1953 respectively. . . . That in order to archive less space in X axis. I drew injuries and fatalities for health, and ,crop and property damages for economic in order to see, in which category was the major hurt, also, I drew the tendencies, for seeing if the hurts for each event have downed or upped.

Data Processing

1. Require to ggplot2.

```
require(ggplot2)
```

2. Verify if the file exist, if no exist then download it and reads it.

```
require(ggplot2)
fileUrl <- "https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2"

if(!file.exists("data/repdata-data-StormData.csv.bz2"))
  download.file(
    fileUrl, "data/repdata-data-StormData.csv.bz2",
    method = "auto", mode = "wb"
  )
dataPrin <- read.csv("data/repdata-data-StormData.csv.bz2")
```

3. Get the year of each event and turn in period of three years. For example 1950, 1951, 1952,1953,1954 will 150, 1950, 1950, 1953, 1953. In order to have to plot more clearly.

```
dataPrin$BGN_DATE <- as.Date(dataPrin$BGN_DATE, "%m/%d/%Y")
dataPrin$year3 <- as.numeric(format(dataPrin$BGN_DATE, '%Y'))
dataPrin$year3 <- dataPrin$year3 - (dataPrin$year3 %% 3)
```

4. To know the events most dangerous for this period of time, first filter subset the events from the 2000.

```
eventsLastYears = subset(
  dataPrin[,
  BGN_DATE>=as.Date('2000-01-01')
)
```

5. For the health aggregated the injuries and fatalities per event.

```
nTop <- 3 #This variable defines the number of events in the ranking.
topEventHealth <- aggregate(
  eventsLastYears[,c('INJURIES', 'FATALITIES')],
  by = list(EVTYPE=eventsLastYears$EVTYPE),
  FUN = sum
)
```

6. Order the events for number of injuries and fatalities, then, assign a nivel of danger and transform in factor to define the order in the panel.

```
topEventHealth <- topEventHealth[
  order(-topEventHealth$INJURIES, -topEventHealth$FATALITIES),
  ][1:nTop,]
topEventHealth$levelDanger <- 1:nTop
topEventHealth$EVTYPE <- as.factor(as.character(topEventHealth$EVTYPE))
```

7. With the ranking of the events most dangerous get all data since 1950 of those events.

```
eventForHealth <- dataPrin[
  dataPrin$EVTYPE %in% topEventHealth$EVTYPE,
]
```

8. Sum the fatalities and injuries and after merge the data to get a data with injuries and fatalities differentiated.

```
topTenHealtInjuries <- aggregate(
  eventForHealth[,c('INJURIES')],
  by = list(
    eventForHealth$EVTYPE,
    eventForHealth$year3),#
  FUN = sum
)
names(topTenHealtInjuries)<-c('EVTYPE', 'year3', 'numPersons')
topTenHealtInjuries$typeHurt <- 'Injurie'
topTenHealtFatalities <- aggregate(
  eventForHealth[,c('FATALITIES')],
  by = list(
    EVTYPE=eventForHealth$EVTYPE,
    year3=eventForHealth$year3),#
  FUN = sum
)
names(topTenHealtFatalities)<-c('EVTYPE', 'year3', 'numPersons')
topTenHealtFatalities$typeHurt <- 'Fatality'
```

```
topTenHealt<- rbind(topTenHealtFatalities, topTenHealtInjuries)

topTenHealt <- merge(topTenHealt, topEventHealth[,c('EVTYPE', 'levelDanger')])

topTenHealt$EVTYPE <- factor(topTenHealt$EVTYPE, levels = topEventHealth$EVTYPE)
```

9. For the economic aggregated the iPROPDMG and CROPDMG per event.

```
topEventEconomic <- aggregate(
  eventsLastYears[,c('PROPDMG', 'CROPDMG')],
  by = list(EVTYPE=eventsLastYears$EVTYPE), #, STATE=eventsLastYears$STATE
  FUN = sum
)
```

9. Order the events for number of injuries and fatalities, then, assign a nivel of danger and transform in factor to define the order in the panel.

```
topEventEconomic <- topEventEconomic[
  order(-topEventEconomic$PROPDMG, -topEventEconomic$CROPDMG),
  ][1:nTop,]
topEventEconomic$levelDanger <- 1:nTop
topEventEconomic$EVTYPE <- as.factor(as.character(topEventEconomic$EVTYPE))
```

10. With the ranking of the events most dangerous get all data since 1950 of those events.

```
eventForEconomic <- dataPrin[
  dataPrin$EVTYPE %in% topEventEconomic$EVTYPE,
  ]
```

11. Sum the fatalities and injuries and after merge the data to get a data with injuries and fatalities differentiated.

```
topTenEconomicPropdmg <- aggregate(
  eventForEconomic[,c('PROPDMG')],
  by = list(
    eventForEconomic$EVTYPE,
    eventForEconomic$year3), #
  FUN = sum
)
names(topTenEconomicPropdmg)<-c('EVTYPE', 'year3', 'costDamage')
topTenEconomicPropdmg$typeHurt <- 'Propdmg'
topTenEconomicCropdmg <- aggregate(
  eventForEconomic[,c('CROPDMG')],
  by = list(
    EVTYPE=eventForEconomic$EVTYPE,
    year3=eventForEconomic$year3), #
  FUN = sum
)
names(topTenEconomicCropdmg)<-c('EVTYPE', 'year3', 'costDamage')
topTenEconomicCropdmg$typeHurt <- 'Cropdmg'
```

```
topTenEconomic<- rbind(topTenEconomicCropdmg, topTenEconomicPropdmg)

topTenEconomic <- merge(topTenEconomic, topEventEconomic[,c('EVTYPE', 'levelDanger')])

topTenEconomic$EVTYPE <- factor(topTenEconomic$EVTYPE, levels = topEventEconomic$EVTYPE)
```

Results

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

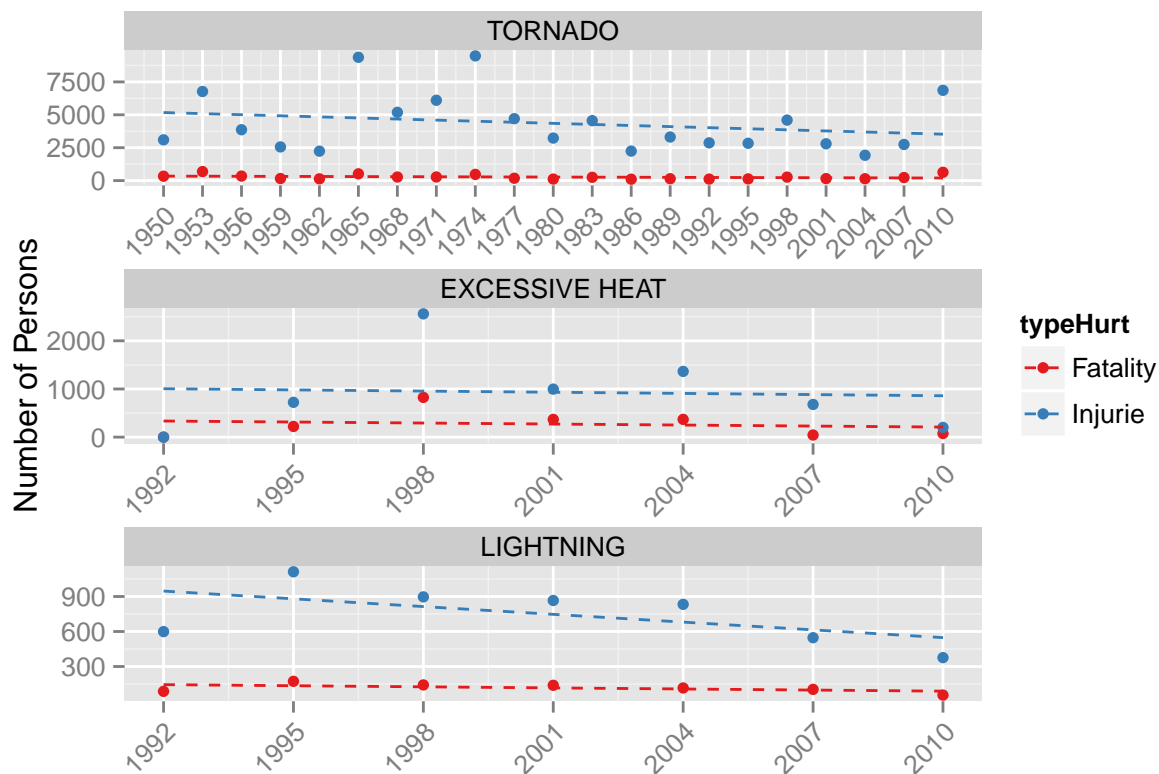
We can see in this panel that the 3 most dangerous event to the health are the Tornado, Excessive Heat and Lightning.

Tornado: although, the tendency have been decreased, in 2010 there was a increase in injuries and fatalities similar to before of 1974.

Excessive heat: although, it appeared until in 1992, have caused an averge of 1000 injuries and 240 fatalities for each three years.

Lightning: it appeared until 1992 and have been decreased.

```
ggplot(
  data = topTenHealt,aes(year3, numPersons, colour = typeHurt))+
  geom_point()+ facet_wrap(~EVTYPE, ncol = 1, scales = "free")+
  guides(col = guide_legend(ncol = 1))+
  theme(axis.text.x = element_text(angle = 45, hjust = 1))+
  scale_x_continuous(breaks=seq(from = 1950,by = 3,to = 2011))+
  geom_smooth(method="lm", se = F, lty = 2)+
  scale_color_brewer(palette="Set1")+ylab('Number of Persons')+xlab('')
```



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

We can see in this panel that the 3 most dangerous event to the health are the Tornado, Excessive Heat and Lightning.

FLASH FLOOD: it appeared until 1992, however, between 2007 and 2010 caused hurt per 5 million of dollar. Has a tendency to increase.

Excessive heat: Has record since 1950, but, of caused hurts per 500.000 dollar to increase to 2'500.000 dollar with a tendency clearly of increase.

Lightning: only appear in 1992, 1995 and 2004 with minimum hurts, but, in 2007 and 2010 had a repently and great increase.

```
ggplot(
  data = topTenEconomic, aes(year3, costDamage, colour = typeHurt)) +
  geom_point() + facet_wrap(~EVTYPE, ncol = 1, scales = "free") +
  guides(col = guide_legend(ncol = 1)) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  scale_x_continuous(breaks = seq(from = 1950, by = 3, to = 2011)) +
  geom_smooth(method = "lm", se = F, lty = 2) +
  scale_color_brewer(palette = "Set1") + ylab('Cost') + xlab('')
```

