NOAAStorm

Fabien

Saturday, January 17, 2015

Health and Economy impact of weather Events in the US during 1950-2011 period

Synopsis

The analysis of weather event database from the National Oceanic & Atmospheric Administration reveals that Flood create the most impact to health and life of US citizens. The Tornadoes are taking the most economically.

Data Processing

first loading the data

```
# if (!file.exists("./Storm.csv.bz2")) {
#         download.file(url="https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2", "./Storm.csv.bz2")
# }
st <- read.table(file = "./Storm.csv.bz2", sep = ",", header = TRUE, stringsAsFactors = FALSE)
```

Then, evaluate the data that need cleaning and repare it.

```
#opts_chunk$set(echo=TRUE,)
```

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
##
## The following object is masked from 'package:stats':
##
##     filter
##
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(tidyr)
# the fields PROPDMGEXP and CROPDMGEXP need to be cleaned.
exp_transform <- function(e) {
        # h -> hundred, k -> thousand, m -> million, b -> billion

        if (e %in% c('h', 'H'))
                return(2)
        else if (e %in% c('k', 'K'))
                return(3)
        else if (e %in% c('m', 'M'))
                return(6)
        else if (e %in% c('b', 'B'))
                return(9)
        else if (!is.na(as.numeric(e))) # if a digit
                return(as.numeric(e))
        else if (e %in% c('', '-', '?', '+'))
                return(0)
        else {
                stop("Invalid exponent value.")
        }
}

st$PROPDMGUSD <- sapply(st$PROPDMGEXP, FUN = exp_transform)
st$CROPDMGUSD <- sapply(st$CROPDMGEXP, FUN = exp_transform)

# date filed transform
st$BGN_DATE <- as.Date(strptime(st[,"BGN_DATE"], format ="%m/%d/%Y %H:%M:%
S"))
st$BGN_year <- as.POSIXlt(st$BGN_DATE)$year

# the event types are quite messy (read unique(st$EVTYPE)), this work will r
equire the input of someone knowledgeable about weather type classification.

# Preparing the data frame for  Fatalities and Injuries

st2 <- st %>% group_by(EVTYPE) %>% summarise(Fatalities = sum(FATALITIES), I
njuries = sum(INJURIES))
st2 <- st2 %>% mutate(rkFat=min_rank(desc(Fatalities)), rkInj=min_rank(desc
(Injuries))) %>%
        gather(Vartotal, total, Fatalities:Injuries) %>% filter(rkFat<=10 |
rkInj<=10) %>%
        droplevels() %>% mutate(EVTYPE = reorder(EVTYPE, rkFat))

# Again Preparing the data frame for  Fatalities and Injuries and showing va
riability between years
st2rk <- st2 %>% select(EVTYPE, rkFat, rkInj)
st21 <- st %>% group_by(EVTYPE, BGN_year) %>%
        summarise(Fatalities = sum(FATALITIES), Injuries = sum(INJURIES))
st21 <- st21 %>% left_join(st2rk) %>%
        gather(Vartotal, total, Fatalities:Injuries) %>% filter(rkFat<=10 |
rkInj<=10) %>%
```

```
        droplevels() %>% mutate(EVTYPE = reorder(EVTYPE, rkFat, mean, na.rm=
TRUE))
```

```
## Joining by: "EVTYPE"
```

```
# Preparing the dataframe for Crop and Properties

st <- st %>% filter(is.numeric(PROPDMG)) %>% mutate(PROPDMGUSD = PROPDMG * 1
0^PROPDMGUSD)
st <- st %>% filter(is.numeric(CROPDMG)) %>% mutate(CROPDMGUSD = CROPDMG * 1
0^CROPDMGUSD)
st3 <- st %>% group_by(EVTYPE) %>% summarise(Property = sum(PROPDMGUSD), Cro
p = sum(CROPDMGUSD), PrCr = Property + Crop)
st3 <- st3 %>% mutate(rkProp=min_rank(desc(Property)), rkCrop=min_rank(desc
(Crop)), rkPrCr=min_rank(desc(PrCr))) %>%
        gather(Vartotal, total, Property:Crop) %>% filter(rkPrCr<=15) %>%
        droplevels() %>% mutate(EVTYPE = reorder(EVTYPE, rkPrCr))
```

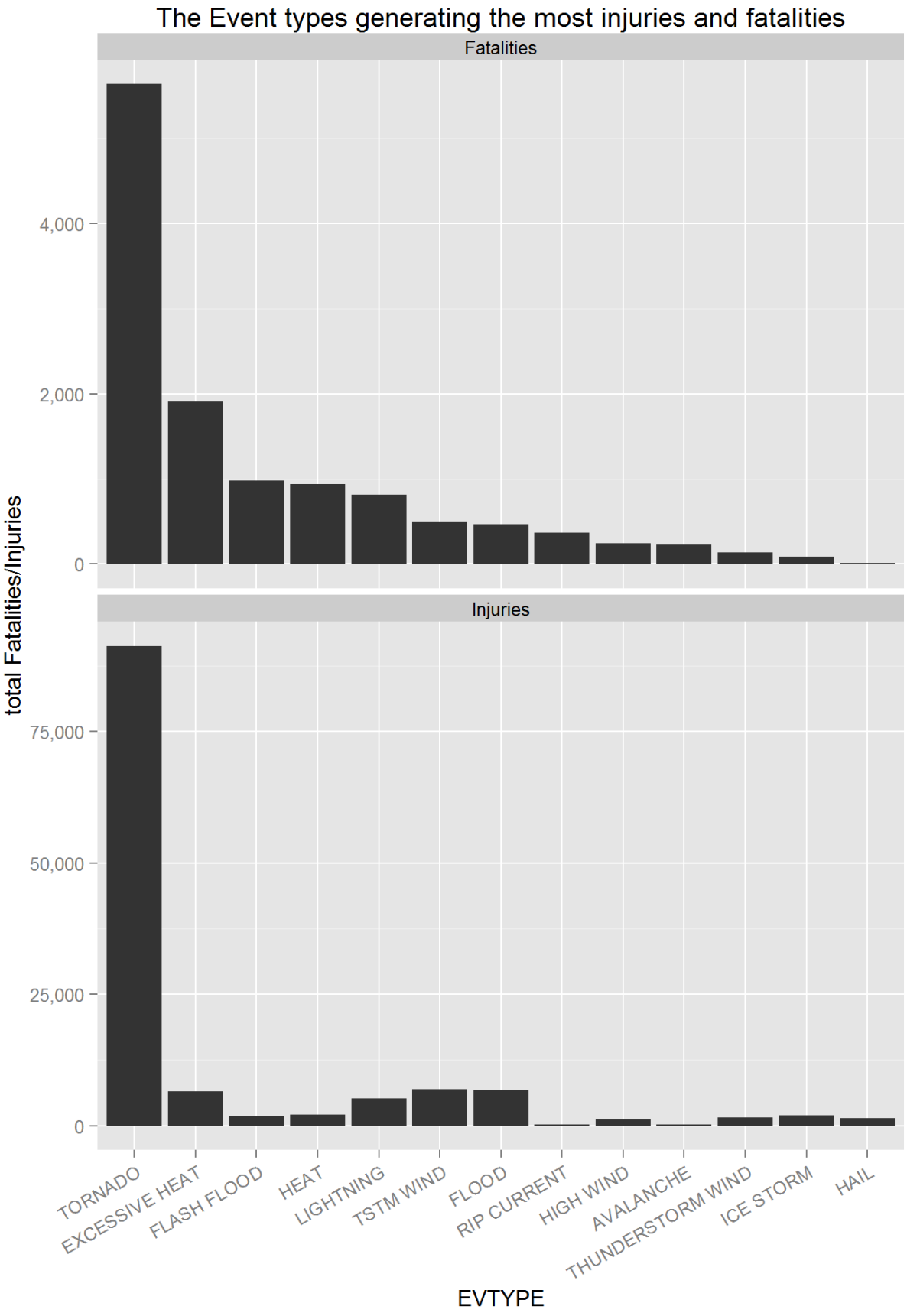
Result

The most harmful events with respect to population health across the United States

the chart below displays the top 15 event types for injuries and fatalities. The one having accounted for the most fatalities are listed first.

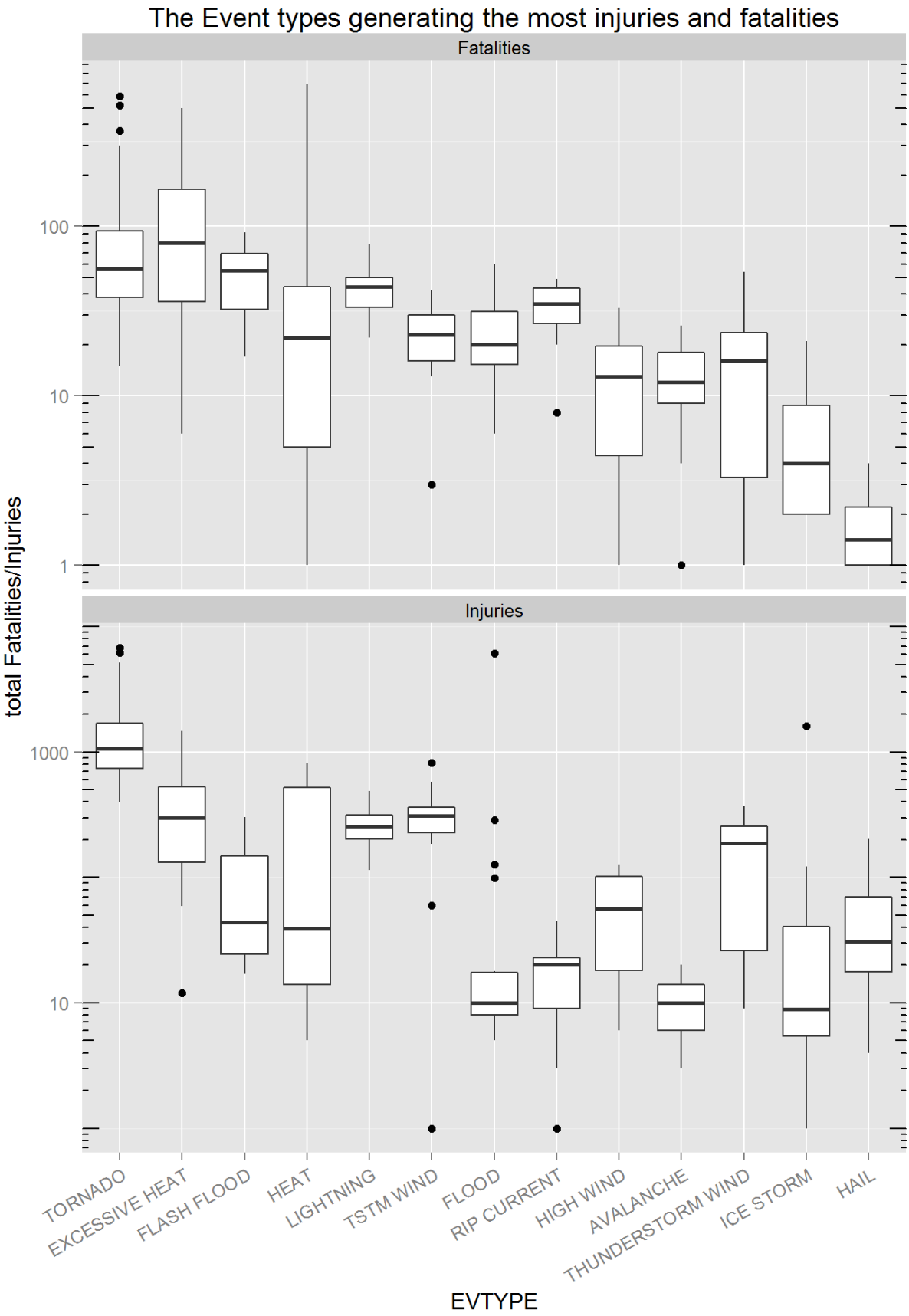
```
library(ggplot2)
library(scales)

ggplot(st2, aes(x=EVTYPE, y= total)) + geom_bar(stat="identity") +
        ylab("total Fatalities/Injuries") + facet_wrap(ncol = 1, ~ Vartota
l, scales="free_y") +
        ggtitle("The Event types generating the most injuries and fatalitie
s") +
        theme(axis.text.x = element_text(angle = 30, hjust = 1)) +
        scale_y_continuous(labels = comma)
```



Tornadoes and Flood have outliers.

```
ggplot(st21, aes(x=EVTYPE, y= total)) + geom_boxplot() +
        scale_y_log10("total Fatalities/Injuries") +
        facet_wrap(ncol = 1, ~ Vartotal, scales="free_y") +
        ggtitle("The Event types generating the most injuries and fatalitie
s") +
        theme(axis.text.x = element_text(angle = 30, hjust = 1)) +
        annotation_logticks(sides = "lr")
```


The Event types generating the most injuries and fatalities
Fatalities
Injuries
total Fatalities/Injuries
100
10
1
1000
10
TORNADO
EXCESSIVE HEAT
FLASH FLOOD
HEAT
LIGHTNING
TSTM WIND
FLOOD
RIP CURRENT
HIGH WIND
AVALANCHE
THUNDERSTORM WIND
ICE STORM
HAIL
EVTYPE

The greatest economic consequences per event type across the United States

The chart below displays the top 10 event types for cumulated crop and property economic consequences.

```
ggplot(st3, aes(x=EVTYPE, y= total)) + geom_bar(stat="identity") +
        ylab("USD damages") + facet_wrap(ncol = 1, ~ Vartotal, scales="free_
y") +
        ggtitle("The Event types generating the most Property and Crop damag
es") +
        theme(axis.text.x = element_text(angle = 30, hjust = 1)) +
        scale_y_continuous(labels = comma)
```