Severe weather events impact on the US (1950-2011)

Patrick Sanzo

Monday, December 15, 2014

Synopsis

This short analysis tries to find the question as to which weather event has the most impact in the US, according to data from the U.S. National Oceanic and Atmospheric Administration's (NOAA) storm database.

To answer this question, we try to focus on two aspects:

- Which types of events are most harmful with respect to population health?
- Across the United States, which types of events have the greatest economic consequences?

We assume the answer to the first question can be found by looking at the TYPE field related to the FATALITIES and INJURIES fields. The second question (economic consequences) could be answered in terms of TYPE against PROPDMG(property damage) and CROPDMG(crop damage) fields.

Further info on NOAA db can be obtained here and here

Data Processing

Assuming we're in the same folder as the compressed data is, we can begin with:

```
stormData <-  read.csv("repdata-data-StormData.csv.bz2")
```

We default NA values to 0, for better arithmetic handling.

```
naValuesFatalities <- is.na(stormData$FATALITIES)
if (any(naValuesFatalities)){
  stormData[naValuesFatalities, ]$FATALITIES <- 0
}

naValuesInjuries <- is.na(stormData$INJURIES)
if (any(naValuesInjuries)){
  stormData[naValuesInjuries,]$INJURIES <- 0
}
```

Let's double check integer values are treated as such

```
stormData$INJURIES <- as.numeric(as.character(stormData$INJURIES))
stormData$FATALITIES <- as.numeric(as.character(stormData$FATALITIES))
```

We create a vector of fatalities and injuries, based on the event type

```
fatalities <- tapply(as.numeric(as.character(stormData$FATALITIES)), stormData$EVTYPE, sum)
injuries <- tapply(as.numeric(as.character(stormData$INJURIES)), stormData$EVTYPE, sum, simplify = T  )
propertyDamage <- tapply(as.numeric(as.character(stormData$PROPDMG)), stormData$EVTYPE, sum)
```

We create data frames for simpler plotting

```
dfFatalities <- data.frame(
  x=names(fatalities[!is.na(fatalities)]),
  y=fatalities[!is.na(fatalities)])

dfInjuries <- data.frame(
  x=names(injuries[!is.na(injuries)]),
  y=injuries[!is.na(injuries)])

dfEconomicConsequences <- data.frame(
  x=names(propertyDamage[!is.na(propertyDamage)]),
  y=propertyDamage[!is.na(propertyDamage)])
```

Results

Assuming we have the ggplot2 library installed (install.packages(ggplot2)), we import the library.

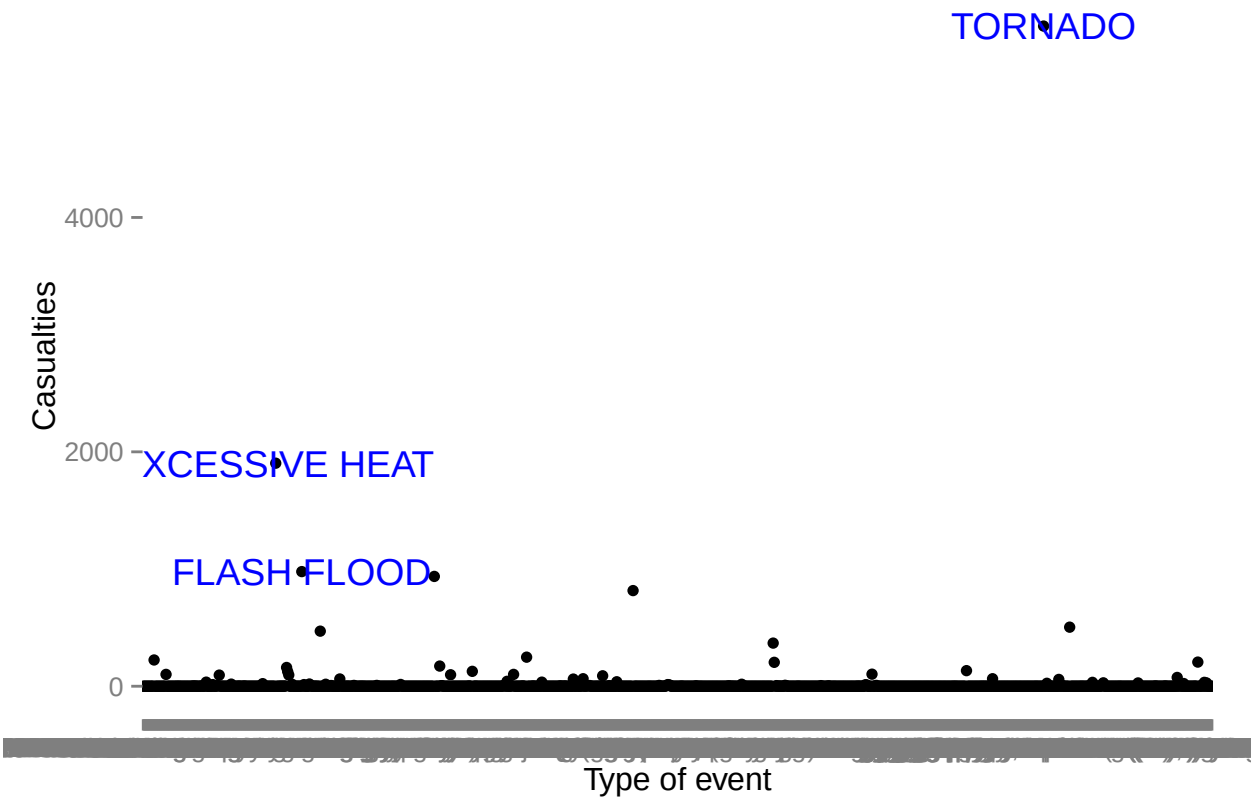
```
library(ggplot2)
```

Most harmful type of events with respect to population health We select the three biggest casualty causes

```
sortedValuesCasualties <- dfFatalities[sort.list(as.numeric(as.character(dfFatalities$y)), decreasing =
```

Let's plot it

```
ggplot(dfFatalities, aes(x,y)) +
  geom_point() + xlab("Type of event") + ylab("Casualties") +
  geom_text(data=sortedValuesCasualties, aes(label=x), col="Blue")
```

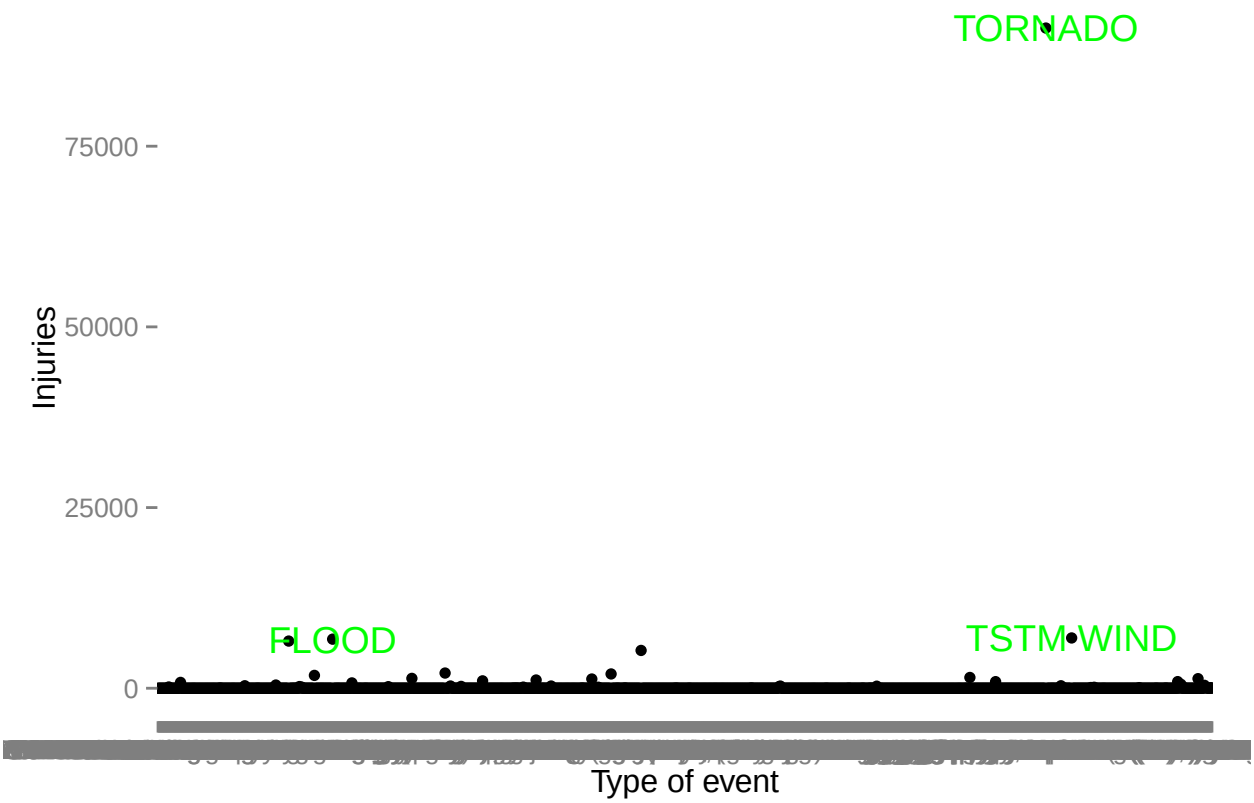


We select the three biggest injury causes

```
sortedValuesInjuries <- dfInjuries[sort.list(as.numeric(as.character(dfInjuries$y)), decreasing = T)[1:3
```

Let's plot it

```
ggplot(dfInjuries, aes(x,y)) +
  geom_point() + xlab("Type of event") + ylab("Injuries") +
  geom_text(data=sortedValuesInjuries, aes(label=x), col="Green")
```



As we can see in both plots, there is an event which greatly surpasses all other events.

Which one is it?

Let's see

```
biggestFatalityCauser <- dfFatalities[dfFatalities$y == max(dfFatalities$y),][1]
biggestInjuryCauser <- dfInjuries[dfInjuries$y == max(dfInjuries$y),][1]
biggestFatalityCauser
```

```
##                x
## TORNADO TORNADO
```

```
biggestInjuryCauser
```

```
##                x
## TORNADO TORNADO
```

Tornadoes are the most harmful events with respect to population health, since they cause a lot more casualties and injuries than any other event. Far from tornadoes, but still significant (related to injuries) are "floods" and "TSTM Wind", and "excessive heat" and "flash flood" (related to casualties).

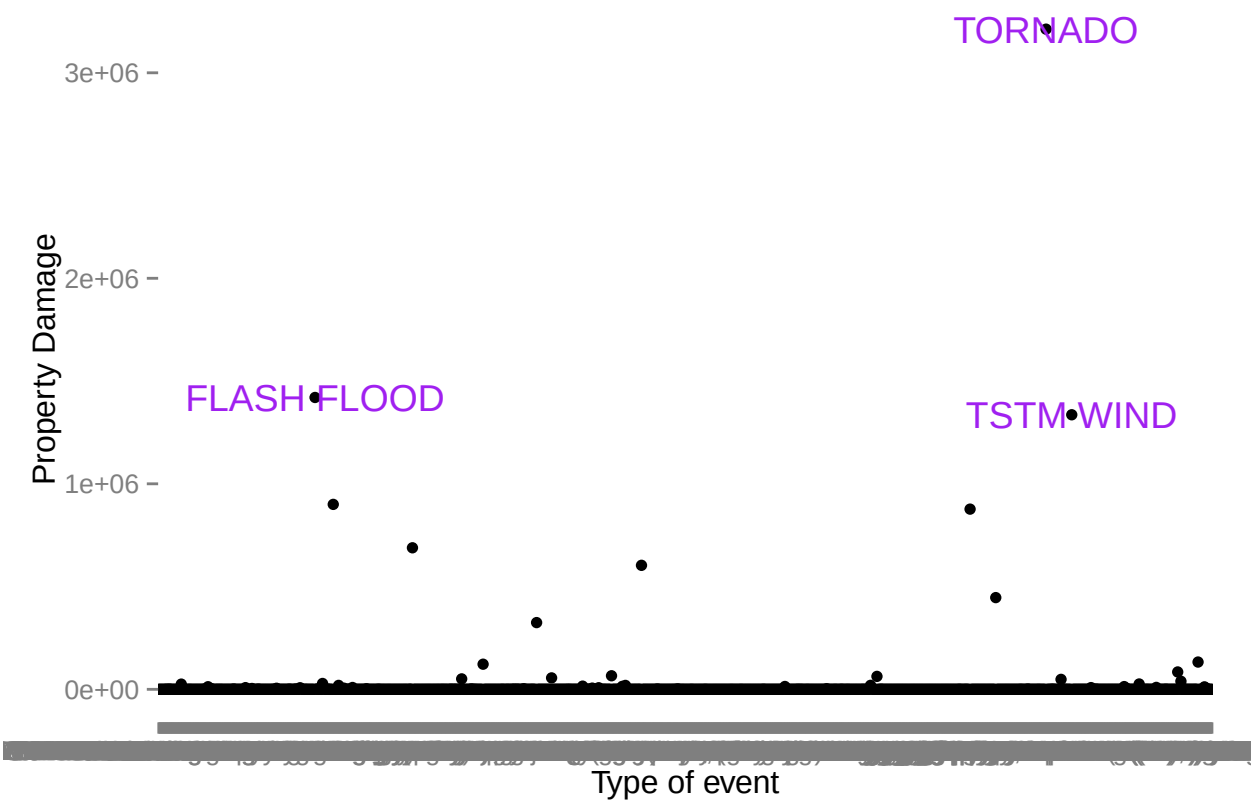
Types of events with the greatest economic consequences (US-wide) Let's see if we can get the event causing the greatest economic consequences. We'll observe the PROPDMG variable against the EVENT TYPE.

We select the three biggest injury causes

```
sortedValuesEconomicConsequences <- dfEconomicConsequences[sort.list(as.numeric(as.character(dfEconomic
```

Let's plot it.

```
ggplot(dfEconomicConsequences, aes(x,y)) +
  geom_point() + xlab("Type of event") + ylab("Property Damage") +
  geom_text(data=sortedValuesEconomicConsequences, aes(label=x), col="Purple")
```




```
biggestCauseOfPropertyDamage <- propertyDamage[propertyDamage==max(propertyDamage)]
biggestCauseOfPropertyDamage
```

```
## TORNADO
## 3212258
```

Again, we can see that **tornadoes have the the greatest economic consequences.**, followed by "flash flood" and "TSTM Wind".