

# Weather event analysis

## Synopsis

This analysis explores the U.S. National Oceanic and Atmospheric Administration's storm database. The goal of the analysis is to find out which weather events have the greatest impact on public health and on the economy.

## Data Processing

First the database is loaded into an R data frame. ##### Download the data

```
filePathRaw <- paste0(getwd(), '/repdata-data-StormData.csv.bz2')
if(!file.exists(filePathRaw))
{
  download.file(url,filePathRaw,method="curl")
}
data <- read.csv(filePathRaw)
```

### Transform event types

Since the weather events are rather 'messy' stored in the database an effort is made to correct the event types.

```
require(stringr)
```

```
## Loading required package: stringr
```

```

data$EVTYPE <- str_trim(tolower(as.character(data$EVTYPE)))
data$EVTYPE <- gsub("^winter weathermix$", "winter weather", data$EVTYPE)
data$EVTYPE <- gsub("^flash flooding$", "flash flood", data$EVTYPE)
data$EVTYPE <- gsub("^urbansml stream fld$", "flood", data$EVTYPE)
data$EVTYPE <- gsub("^avalance$|^landslide$", "avalanche", data$EVTYPE)
data$EVTYPE <- gsub("^extreme coldwind chill$|^extreme cold$", "extreme cold/wind chill", data$EVTYPE)
data$EVTYPE <- gsub("winds|windss|wnd", "wind", data$EVTYPE)
data$EVTYPE <- gsub("tstm|thunderstorm", "thunderstorm", data$EVTYPE)
data$EVTYPE <- gsub("floods|flooding|fld", "flood", data$EVTYPE)
data$EVTYPE <- gsub("funnels", "funnel", data$EVTYPE)
data$EVTYPE <- gsub("hailstorm|hailstorms", "hail", data$EVTYPE)
data$EVTYPE <- gsub(".*surf.*", "high surf", data$EVTYPE)
data$EVTYPE <- gsub("winterymix|wintrymix|wintery|wintry", "winter", data$EVTYPE)
data$EVTYPE <- gsub("rural", "", data$EVTYPE)
data$EVTYPE <- gsub("thunderstorm wind.*", "thunderstorm wind", data$EVTYPE)
data$EVTYPE <- gsub("high wind.*", "high wind", data$EVTYPE)
data$EVTYPE <- gsub("^wildforest fire$", "wildfire", data$EVTYPE)
data$EVTYPE <- gsub("^frostfreeze$", "frost/freeze", data$EVTYPE)
data$EVTYPE <- gsub("^lakeeffect snow$", "lake-effect snow", data$EVTYPE)
data$EVTYPE <- gsub("^snow$", "heavy snow", data$EVTYPE)
data$EVTYPE <- gsub("^coldwind chill$", "cold/wind chill", data$EVTYPE)
data$EVTYPE <- gsub("^fog$", "freezing fog", data$EVTYPE)

```

## Transform economic data

The economic data is divided into a number and a quantity (like 'b' billion). The quantities are used to convert the economic data into 1 number per weather event instance.

```

data$PROPDMGEXP <- toupper(data$PROPDMGEXP)
data$PROPDMG <- ifelse(data$PROPDMGEXP == "B", data$PROPDMG * 1e+09, ifelse(data$PROPDMGEXP == "M", data$PROPDMG * 1e+06, ifelse(data$PROPDMGEXP == "K", data$PROPDMG * 1000, ifelse(data$PROPDMGEXP == "H", data$PROPDMG * 100, data$PROPDMG))))

data$CROPDMGEXP <- toupper(data$CROPDMGEXP)
data$CROPDMG <- ifelse(data$CROPDMGEXP == "B", data$CROPDMG * 1e+09, ifelse(data$CROPDMGEXP == "M", data$CROPDMG * 1e+06, ifelse(data$CROPDMGEXP == "K", data$CROPDMG * 1000, ifelse(data$CROPDMGEXP == "H", data$CROPDMG * 100, data$CROPDMG))))

```

## Prepare health data

The health data (injuries / fatalities) is summarized, ordered and the top-10 is picked.

```

eventFatalities <- aggregate(data$FATALITIES, by=list(data$EVTYPE), FUN=sum)
colnames(eventFatalities) <- c("event", "fatalities")

eventInjuries <- aggregate(data$INJURIES, by=list(data$EVTYPE), FUN=sum)
colnames(eventInjuries) <- c("event", "injuries")

eventHealth <- merge(eventInjuries, eventFatalities, by="event")
eventHealth <- eventHealth[order(eventHealth$injuries, decreasing=T),]
eventHealth <- eventHealth[1:10,]

```

## Prepare economic data

The economic data (property / crop damage) is summarized, ordered and the top-10 is picked.

```

eventPropertyDamages <- aggregate(data$PROPDMG, by=list(data$EVTYPE), FUN=sum)
colnames(eventPropertyDamages) <- c("event", "property_damage")

eventCropDamages <- aggregate(data$CROPDMG, by=list(data$EVTYPE), FUN=sum)
colnames(eventCropDamages) <- c("event", "crop_damage")

eventDamages <- merge(eventPropertyDamages, eventCropDamages, by="event")
eventDamages$total <- eventDamages$property_damage + eventDamages$crop_damage
eventDamages <- eventDamages[order(eventDamages$total, decreasing=T),]
row.names(eventDamages) <- NULL

plotData <- eventDamages[1:10,]
plotData$total <- NULL

```

# Results

## Health

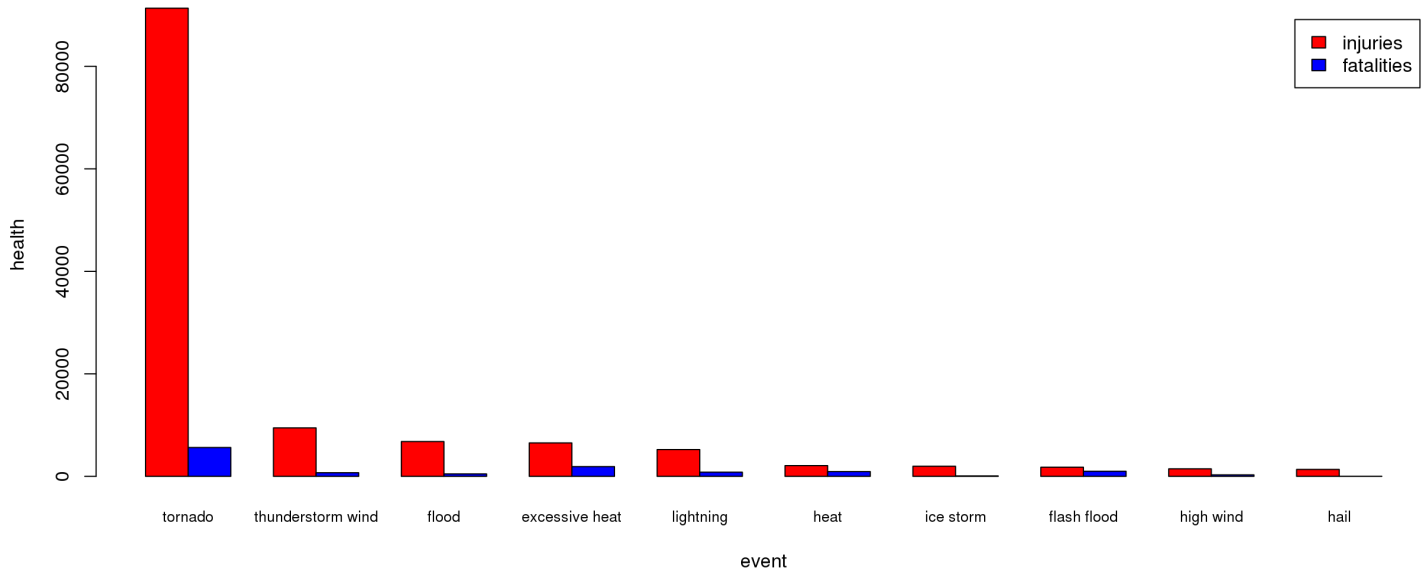
Across the United States, which types of events are most harmful with respect to population health?

```

barplot(t(eventHealth[-1]), names.arg=eventHealth$event, col=c("red", "blue"), beside=T, legend
=colnames(eventHealth[-1]), xlab="event", ylab="health", main="Injuries / fatalities per event
(top 10)", cex.name=0.8)

```

Injuries / fatalities per event (top 10)



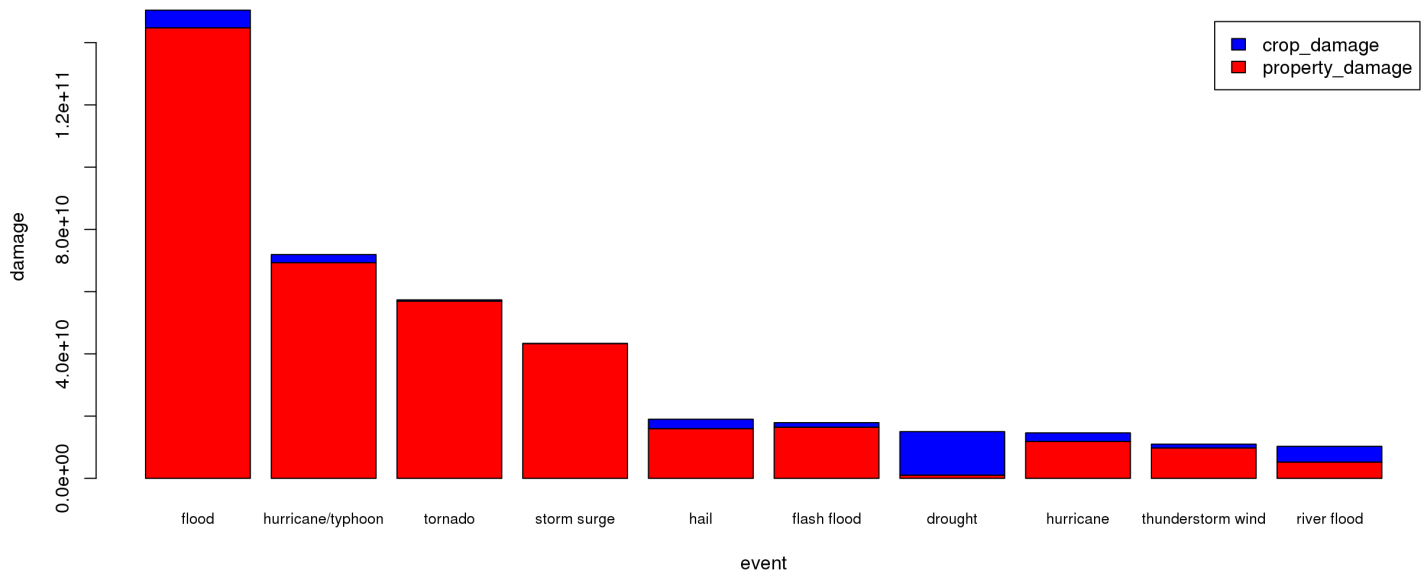
The figure shows the top-10 events with the number of fatalities and injuries. The weather event Tornado is most harmful to population health.

## Economic damage

Across the United States, which types of events have the greatest economic consequences?

```
barplot(t(plotData[-1]), names.arg=plotData$event, col=c("red", "blue"), legend=colnames(plotData[-1]), xlab="event", ylab="damage", main="Economic damage per event (top 10)", cex.name=0.8)
```

Economic damage per event (top 10)



The figure shows the top-10 events with the total damage. The weather event Flood has the greatest impact on the US economy.