

# Impact of natural events on health and economics in the US

The following analysis makes use of the NOAA database to identify and explore the impact of the most severe weather events in the US. The impact of events on public health is measured in terms of injuries and fatalities. The economic damage is measured by a subset of the whole dataset that contains full information about the absolute damage in dollars.

## Data Processing

loading relevant packages and unzipping the data file

```
library(R.utils)
library(data.table)
library(dplyr)
library(ggplot2)
library(ggthemes)
```

reading in the data and convert to data.table

```
alldata <- read.csv("repdata-data-StormData.csv")
alldata <- as.data.table(alldata)
```

summarizing the sum of injuries and fatalities per natural event in a new df

```
healthdmg <- alldata %>%
  select(c(EVTYPE, INJURIES, FATALITIES)) %>%
  group_by(EVTYPE) %>%
  summarise(
    inj <- sum(INJURIES),
    fat <- sum(FATALITIES))

names(healthdmg) <- c("event", "inj", "fat")
```

create two data frames, where the damage of events to public health is ordered for 1) injuries and 2) fatalities. Subsequently, assign the event as ordered factor to show them in the right order in ggplot

```
inj <- arrange(healthdmg, desc(inj))
fat <- arrange(healthdmg, desc(fat))

inj$event <- factor(inj$event, levels = inj$event, ordered = T)
fat$event <- factor(fat$event, levels = fat$event, ordered = T)
```

exploring the economic damage of weather events

having a closer look on the variables that define the magnitude of damage in dollars

```
table(alldata$PROPDMGEXP)
```

```
##
##      -      ?      +      0      1      2      3      4      5
## 465934    1    8    5   216   25   13    4    4   28
##      6      7      8      B      h      H      K      m      M
##      4      5      1     40      1      6 424665    7 11330
```

```
table(alldata$CROPDMGEXP)
```

```
##
##           ?         0         2         B         k         K         m         M
## 618413      7        19        1        9        21 281832        1       1994
```

put them all in lower case to combine the important elements

```
alldata$PROPDGMGEXP <- tolower(alldata$PROPDGMGEXP)
alldata$CROPDMGEXP <- tolower(alldata$CROPDMGEXP)
```

now map the letters to actual multipliers

```
objectdmg <- alldata %>%
  filter((PROPDGMGEXP %in% c("h", "k", "m", "b") &
          CROPDMGEXP %in% c("h", "k", "m", "b"))) %>%
  select(c(EVTYPE, PROPDMG, PROPDGMGEXP, CROPDMG, CROPDMGEXP))

get_mult <- function(x) {
  if (x == "h") x <- 100
  if (x == "k") x <- 1000
  if (x == "m") x <- 1000000
  if (x == "b") x <- 1000000000
  x
}

objectdmg$PROPDGMGEXP <- sapply(objectdmg$PROPDGMGEXP, get_mult)
objectdmg$CROPDMGEXP <- sapply(objectdmg$CROPDMGEXP, get_mult)
```

we don't want to use the exponential notation!

```
options(scipen = 2)
```

now, create a new data frame containing the combined damage from crop and property to get insights into the full economic impact of each event

```
dmg <- objectdmg %>%
  mutate(DAMAGE = PROPDMG * PROPDGMGEXP + CROPDMG * CROPDMGEXP) %>%
  group_by(EVTYPE) %>%
  summarise(dmg = (sum(DAMAGE)/1000000)) %>%
  arrange(desc(dmg))

names(dmg) <- c("event", "dmg")
```

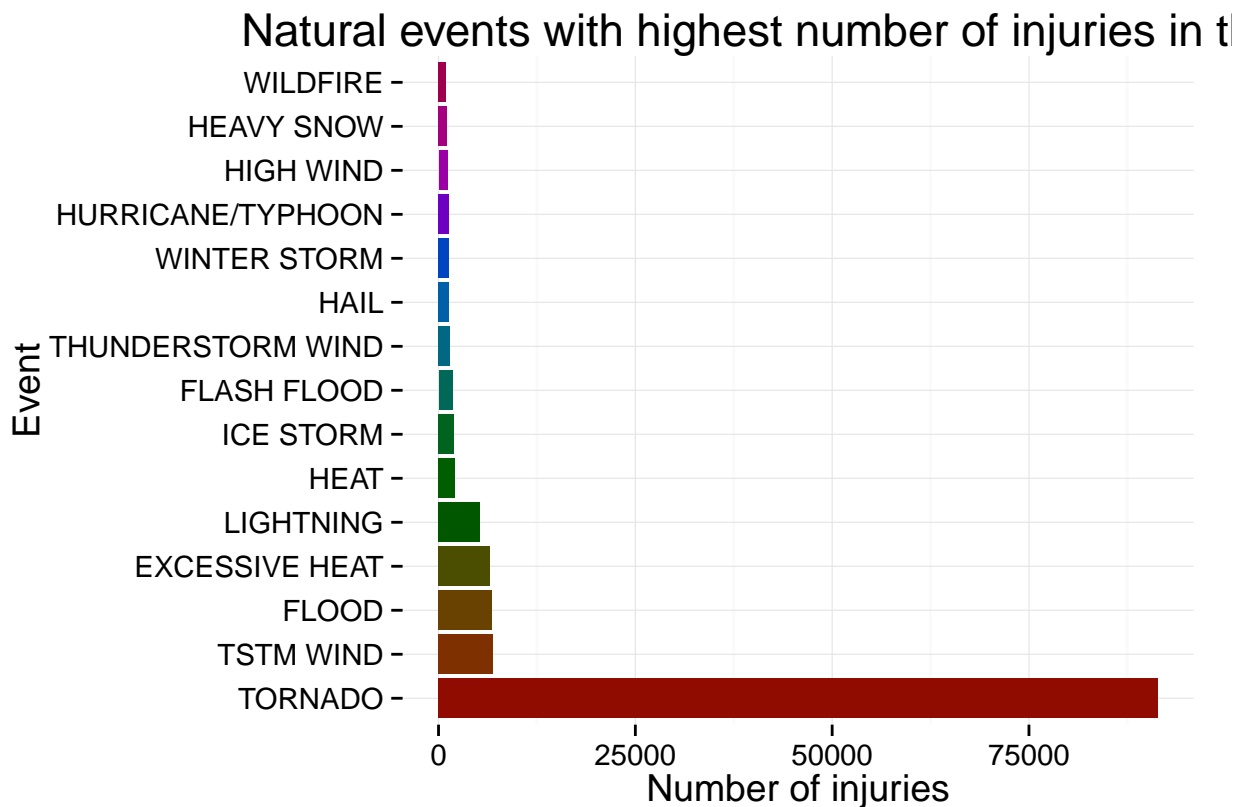
and create an ordered factor again to plot them in descending order

```
dmg$event <- factor(dmg$event, levels = dmg$dmg, ordered = T)
```

## Results

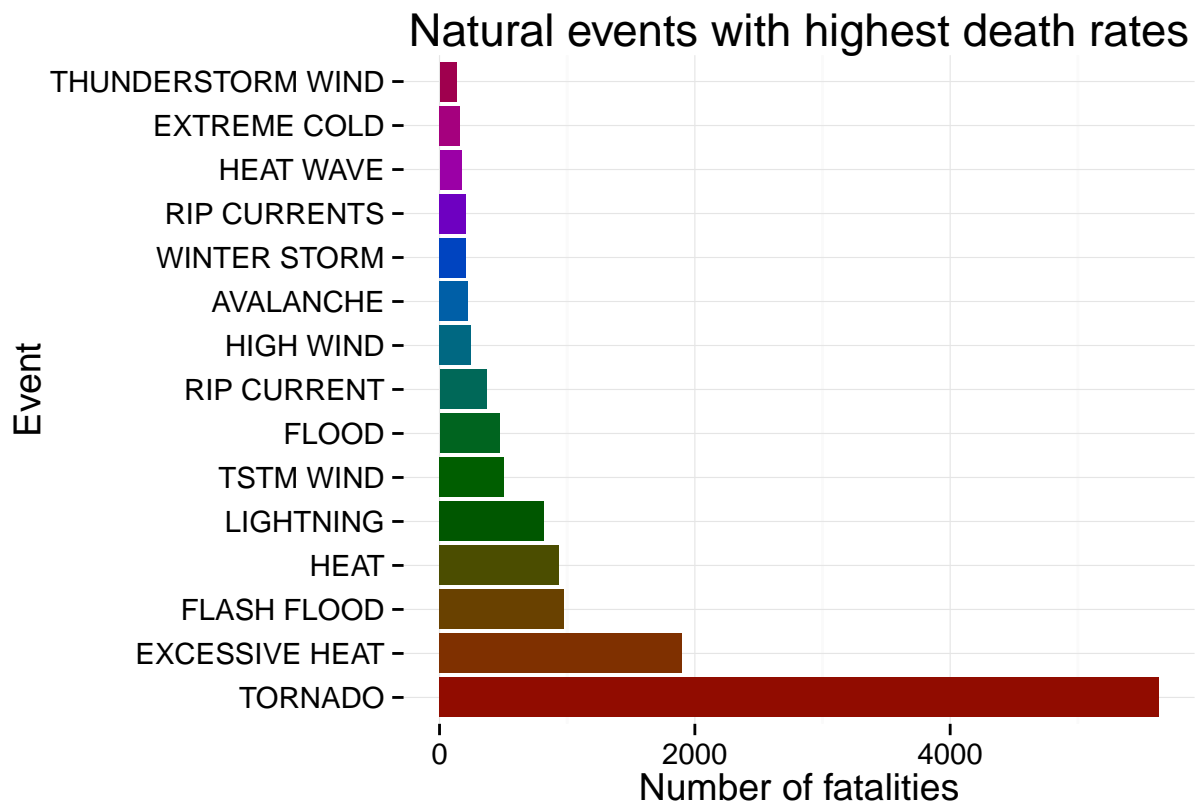
plotting the most severe events concerning human injuries

```
ggplot(inj[1:15, ], aes(x = event, y = inj, fill = event)) +
  geom_bar(stat = "identity") +
  scale_fill_hue(l=30) +
  coord_flip() +
  xlab("Event") +
  ylab("Number of injuries") +
  theme_minimal(base_size = 14) +
  guides(fill=FALSE) +
  ggtitle("Natural events with highest number of injuries in the US")
```



and fatalities

```
ggplot(fat[1:15, ], aes(x = event, y = fat, fill = event)) +
  geom_bar(stat = "identity") +
  scale_fill_hue(l=30) +
  coord_flip() +
  xlab("Event") +
  ylab("Number of fatalities") +
  theme_minimal(base_size = 14) +
  guides(fill=FALSE) +
  ggtitle("Natural events with highest death rates")
```



And, last but not least, the weather events with the highest economic impact, based on property and crop damage

```
ggplot(dmg[1:15, ], aes(x = event, y = dmg, fill = event)) +
  geom_bar(stat = "identity") +
  scale_fill_hue(l=30) +
  coord_flip() +
  xlab("Event") +
  ylab("Damage in Million Dollars") +
  theme_minimal(base_size = 14) +
  guides(fill=FALSE) +
  ggtitle("Economic impact of weather events in the US")
```

