Reproducible Research Project 2: NOAA Storm Data Analysis

Synopsis

An analysis of NOAA Storm Events Data ranging from 1950 to 2011. We aggregate the data and look at the total number of injuries, fatalities, and amount of damage caused. Overall, floods are responsible for the most economic damage, but tornadoes cause the most injuries and fatalities. They are also the 3rd leading cause of damage.

Data Processing

Data is downloaded from a url (https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2) and unzipped using the bunzip2 function of the R.utils library.

```
# library('R.utils')
#
download.file('https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2','NOAA.csv.bz2',r

# bunzip2('NOAA.csv.bz2')
data <- read.csv("NOAA.csv", header = TRUE)
library(ggplot2)
```

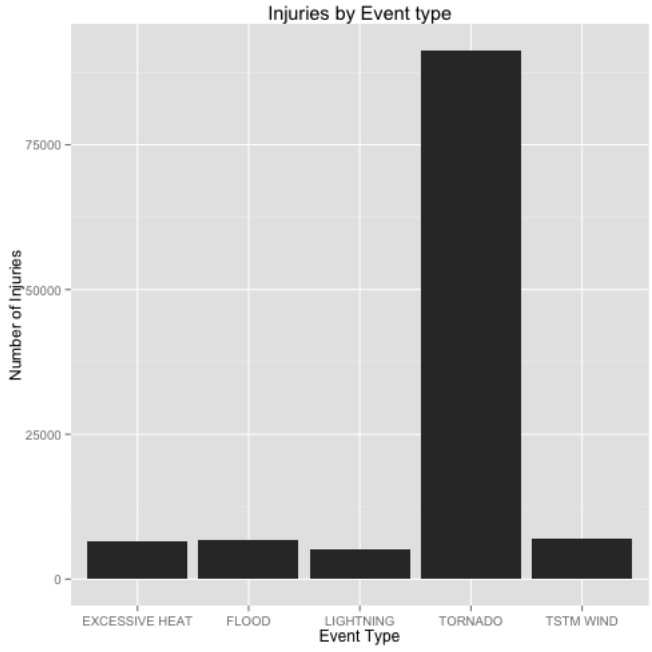
What causes the most injuries?

We can see below that Tornadoes cause the most injuries.

```
injuries <- aggregate(data$INJURIES, by = list(EVTYPE = data$EVTYPE), sum)
injuries <- injuries[order(injuries$x, decreasing = TRUE), ]
head(injuries, 10)
```

```
##                EVTYPE     x
## 834           TORNADO 91346
## 856         TSTM WIND  6957
## 170             FLOOD  6789
## 130    EXCESSIVE HEAT  6525
## 464         LIGHTNING  5230
## 275              HEAT  2100
## 427         ICE STORM  1975
## 153       FLASH FLOOD  1777
## 760 THUNDERSTORM WIND  1488
## 244              HAIL  1361
```

```
ggplot(injuries[1:5, ], aes(EVTYPE, y = x)) + geom_bar(stat = "identity") +
    xlab("Event Type") + ylab("Number of Injuries") + ggtitle("Injuries by Event type")
```



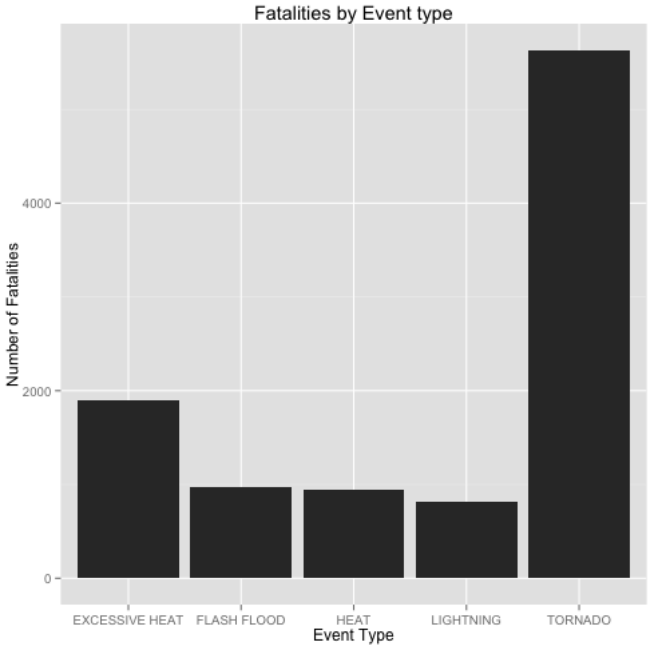
What causes the most fatalities?

We can see that Tornadoes also cause the most fatalities.

```
fatalities <- aggregate(data$FATALITIES, by = list(EVTYPE = data$EVTYPE), sum)
fatalities <- fatalities[order(fatalities$x, decreasing = TRUE), ]
head(fatalities, 10)
```

```
##             EVTYPE    x
## 834        TORNADO 5633
## 130 EXCESSIVE HEAT 1903
## 153    FLASH FLOOD  978
## 275           HEAT  937
## 464      LIGHTNING  816
## 856      TSTM WIND  504
## 170          FLOOD  470
## 585    RIP CURRENT  368
## 359      HIGH WIND  248
## 19       AVALANCHE  224
```

```
ggplot(fatalities[1:5, ], aes(EVTYPE, y = x)) + geom_bar(stat = "identity") +
    xlab("Event Type") + ylab("Number of Fatalities") + ggtitle("Fatalities by Event type")
```



Calculating Damage

The dataset provides an exponential multiplier in the form of 10^x. However the data is not consistent. Sometimes we see a 'B' instead of '9' for billion. We also sometimes see a 'b' instead of a capital 'B'. The following code remedies the discrepancies. At the end, since we are most interested in total damage, we combine the values for Crop and Property damage.

```
# First, make everything upper case
data$PROPDMGEXP <- toupper(data$PROPDMGEXP)
data$CROPDMGEXP <- toupper(data$CROPDMGEXP)
unique(c(data$PROPDMGEXP, data$CROPDMGEXP))
```

```
##  [1] "K" "M" ""  "B" "+" "0" "5" "6" "?" "4" "2" "3" "H" "7" "-" "1" "8"
```

```
datasub <- data[, c("EVTYPE", "PROPDMG", "PROPDMGEXP", "CROPDMG", "CROPDMGEXP")]
datasub[datasub$PROPDMGEXP %in% c("", "+", "-", "?"), "PROPDMGEXP"] <- "0"
datasub[datasub$CROPDMGEXP %in% c("", "+", "-", "?"), "CROPDMGEXP"] <- "0"
unique(c(datasub$PROPDMGEXP, datasub$CROPDMGEXP))
```

```
##  [1] "K" "M" "0" "B" "5" "6" "4" "2" "3" "H" "7" "1" "8"
```

```
# Create 10^x substitutions for Billion, Hundred, Kilo, and Million
datasub[datasub$PROPDMGEXP == "B", "PROPDMGEXP"] <- 9
datasub[datasub$CROPDMGEXP == "B", "CROPDMGEXP"] <- 9
datasub[datasub$PROPDMGEXP == "M", "PROPDMGEXP"] <- 6
datasub[datasub$CROPDMGEXP == "M", "CROPDMGEXP"] <- 6
datasub[datasub$PROPDMGEXP == "K", "PROPDMGEXP"] <- 3
datasub[datasub$CROPDMGEXP == "K", "CROPDMGEXP"] <- 3
datasub[datasub$PROPDMGEXP == "H", "PROPDMGEXP"] <- 2
datasub[datasub$CROPDMGEXP == "H", "CROPDMGEXP"] <- 2
unique(c(datasub$PROPDMGEXP, datasub$CROPDMGEXP))
```

```
##  [1] "3" "6" "0" "9" "5" "4" "2" "7" "1" "8"
```

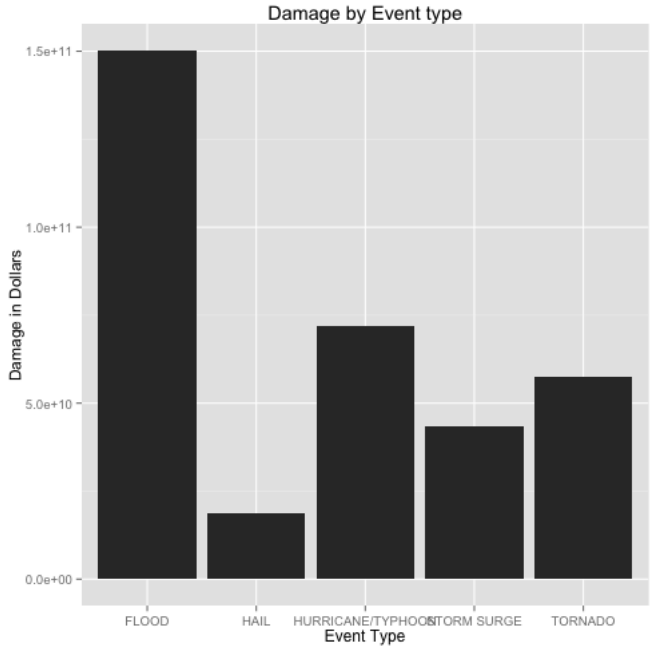
```
# Now combine the exponent with the value
datasub$PROPDMGEXP <- 10^(as.numeric(datasub$PROPDMGEXP))
datasub$CROPDMGEXP <- 10^(as.numeric(datasub$CROPDMGEXP))
datasub[is.na(datasub$PROPDMG), "PROPDMG"] <- 0
datasub[is.na(datasub$CROPDMG), "CROPDMG"] <- 0
# Calculate the total damage
datasub <- within(datasub, TOTALDMG <- PROPDMG * PROPDMGEXP + CROPDMG * CROPDMGEXP)

DamageByType <- aggregate(datasub$TOTALDMG, by = list(EVTYPE = datasub$EVTYPE),
    FUN = sum)
DamageByType <- DamageByType[order(DamageByType$x, decreasing = TRUE), ]
```

We can see that Floods cause the most damage. Floods also cause more than twice as much damage as hurricanes/typhoons, which came in second.

```
# Show the 10 most damaging
head(DamageByType, 10)
```

```
##                EVTYPE         x
## 170             FLOOD 1.503e+11
## 411 HURRICANE/TYPHOON 7.191e+10
## 834           TORNADO 5.736e+10
## 670       STORM SURGE 4.332e+10
## 244              HAIL 1.876e+10
## 153       FLASH FLOOD 1.824e+10
## 95            DROUGHT 1.502e+10
## 402         HURRICANE 1.461e+10
## 590       RIVER FLOOD 1.015e+10
## 427         ICE STORM 8.967e+09
```

```
DamageSubset <- DamageByType[1:5, ]
ggplot(DamageSubset, aes(EVTYPE, y = x)) + geom_bar(stat = "identity") + xlab("Event Type") +
    ylab("Damage in Dollars") + ggtitle("Damage by Event type")
```



Summary

Floods are responsible for the most economic damage, but tornadoes cause the most injuries and fatalities. They are also the 3rd leading cause of damage.