

# Analysis of Storm Data

## Synopsis

The storm data in this analysis is from the U.S. National Oceanic and Atmospheric Administration's (NOAA) storm database. The storm data are recorded starting in the year 1950 and ending in November 2011. In this analysis, the population health is evaluated by the sum of fatalities and injuries, and the economic damage is evaluated by the sum of property damage and crop damage.

The top five event types causing harm to population health are tornado, excessive heat, tstm wind, flood, and lighting.

The top five event types causing economic damages are flood, hurrican/typhoon, tornado, storm surge, and hail.

## Data Processing

Download the data from the course website. The documentations of the data are available at:

- National Weather Service Storm [Data Documentation](#)
- National Climatic Data Center Storm Events [FAQ](#)

```
download.file("https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2", destfile="repdata_data_StormData.csv.bz2")
data <- read.csv("repdata_data_StormData.csv.bz2")
# The types of events:
names(data)
```

```
## [1] "STATE_" "BGN_DATE" "BGN_TIME" "TIME_ZONE" "COUNTY"
## [6] "COUNTYNAME" "STATE" "EVTYPE" "BGN_RANGE" "BGN_AZI"
## [11] "BGN_LOCATI" "END_DATE" "END_TIME" "COUNTY_END" "COUNTYENDN"
## [16] "END_RANGE" "END_AZI" "END_LOCATI" "LENGTH" "WIDTH"
## [21] "F" "MAG" "FATALITIES" "INJURIES" "PROPDMG"
## [26] "PROPDMGEXP" "CROPDMG" "CROPDMGEXP" "WFO" "STATEOFFIC"
## [31] "ZONENAMES" "LATITUDE" "LONGITUDE" "LATITUDE_E" "LONGITUDE_"
## [36] "REMARKS" "REFNUM"
```

## Results

```
library(dplyr)
```

1. Find the types of events that are most harmful with respect to population health across the U.S. The population health is represented by the sum of fatalities and injuries due to the events.

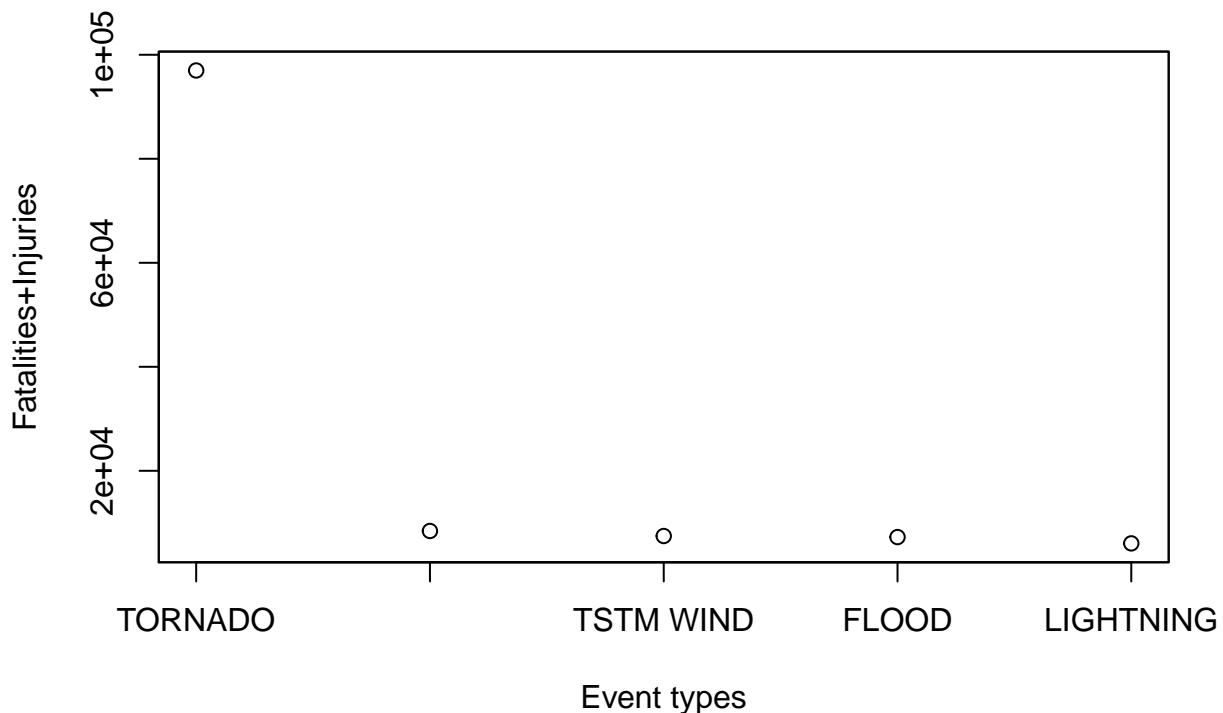
```
##
## Attaching package: 'dplyr'
##
## The following objects are masked from 'package:stats':
##
## filter, lag
##
```

```
## The following objects are masked from 'package:base':
##
## intersect, setdiff, setequal, union
```

```
# subset data
datahealth <- select(data, c(EVTYPE,FATALITIES, INJURIES))
# calculate the population health (pophealth) parameter and subset data again
health <- mutate(datahealth, pophealth=FATALITIES+INJURIES)
health2 <- select(health, c(EVTYPE, pophealth))
# list the top 5 events that are the most harmful to pophealth
grp <- group_by(health2, EVTYPE)
x<- summarize(grp, popfi = sum(pophealth))
x5 <- head(arrange(x,desc(popfi)),5)
x5
```

```
## Source: local data frame [5 x 2]
##
##      EVTYPE popfi
##      (fctr) (dbl)
## 1  TORNADO 96979
## 2 EXCESSIVE HEAT 8428
## 3   TSTM WIND 7461
## 4   FLOOD 7259
## 5 LIGHTNING 6046
```

```
# plot the top 5 events harmful to pophealth
plot(seq(x5$popfi), x5$popfi,xlab="Event types",ylab="Fatalities+Injuries",xaxt="n")
axis(side=1,at=c(1:5),labels=x5$EVTYPE)
box()
```



```

# subset data with economic impact data
dataeco <- select(data, c(EVTYPE,PROPDMG,PROPDMGEXP,CROPDMG,CROPDMGEXP))
eco <- mutate(dataeco, propdmgexp2= (PROPDMGEXP=="K")*10^3+(PROPDMGEXP=="M")*10^6+(PROPDMGEXP=="B")*10^9)
# subset data
eco2 <- select(eco, c(EVTYPE, PROPDMG, propdmgexp2, CROPDMG, cropdmgexp2))
eco3 <- mutate(eco2, propdmgamt=PROPDMG*propdmgexp2, cropdmgamt=CROPDMG*cropdmgexp2)
eco4 <- mutate(eco3, dmgamt=propdmgamt+cropdmgamt)
eco5 <- select(eco4, c(EVTYPE, dmgamt))
# find the top 5 event types that have the greatest economic consequences
ecogrp <- group_by(eco5, EVTYPE)
y<- summarize(ecogrp, totdmgamt = sum(dmgamt))
y5 <- head(arrange(y,desc(totdmgamt)), 5)
y5

```

2. Find the types of events have the greatest economic consequences accorss the U.S.

```

## Source: local data frame [5 x 2]
##
##           EVTYPE      totdmgamt
##           (fctr)      (dbl)
## 1           FLOOD 150319678250
## 2 HURRICANE/TYPHOON 71913712800
## 3           TORNADO 57340613590
## 4     STORM SURGE 43323541000
## 5           HAIL 18752904170

```

```

# plot the top 5 events harmful to pophealth
plot(seq(y5$totdmgamt), y5$totdmgamt,xlab="Event types",ylab="Total Economic Damage Amount (dollars)",
vlabels <- y5$EVTYPE
vlabels <- gsub("/", "", vlabels)
axis(side=1,at=c(1:5),labels=vlabels,cex.axis=0.8)
box()

```

