Risk analysis of the NOAA Storm Database

This is a test analysis for the Coursera course "Reprodcible Research" (Peer Assessment 2).

Synopsis

The basic goal of this assignment is to explore the NOAA Storm Database and answer some basic questions about severe weather events. The data analysis wants to address two main questions: (i) Across the United States, which types of events are most harmful with respect to population health? (ii) Across the United States, which types of events have the greatest economic consequences?

However, this report does not want to make any specific recommendations.

Data processing

The U.S. National Oceanic and Atmospheric Administration's (NOAA) Storm database is downloaded from the public server and read into a table. The column "REMARKS" is removed from further analysis, since it contains free text information that blows up table size and makes displaying cumbersome. The resulting object is stored for faster access.

```
setwd("~/workspace/coursera")
suppressPackageStartupMessages(library(data.table))
suppressPackageStartupMessages(library(gplots))

noaa_link= "http://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2"
noaa_file= "repData_peer2/noaa_stormdata.csv.bz2"
noaa_bak= "repData_peer2/noaa_table.RDS"

if(!file.exists(noaa_bak)){
    download.file(noaa_link, destfile=noaa_file)
    colidx_remarks= 36
    noaa_table= as.data.table(read.csv(noaa_file))[,setdiff(1:37,colidx_remarks), with=F]
    saveRDS(noaa_table, file=noaa_bak)
}else{
    noaa_table= readRDS(noaa_bak)
}
```

In this analysis I want to focus on human, property and crop damage, therefore the table is restricted to the following columns: EVTYPE (weather events), FATALITIES (number of fatalities), INJURIES (number of injuries), PROPDMG (property damage), CROPDMG (crop damage).

```
dmg_noaa_table= noaa_table[, list(EVTYPE, FATALITIES, INJURIES, PROPDMG, CROPDMG)]
dmg_noaa_table
```

```
##            EVTYPE FATALITIES INJURIES PROPDMG CROPDMG
##      1:   TORNADO          0       15    25.0       0
##      2:   TORNADO          0        0     2.5       0
##      3:   TORNADO          0        2    25.0       0
##      4:   TORNADO          0        2     2.5       0
```

```
##      5:    TORNADO           0         2      2.5        0
##    ---
## 902293:  HIGH WIND          0         0      0.0        0
## 902294:  HIGH WIND          0         0      0.0        0
## 902295:  HIGH WIND          0         0      0.0        0
## 902296:   BLIZZARD          0         0      0.0        0
## 902297: HEAVY SNOW          0         0      0.0        0
```

Generate basic statistics about the data. Which events are documented? How frequent are they? What is distribution of all damages, i.e. uniform or normally distributed?

```
num_evtypes=unique(dmg_noaa_table$EVTYPE)
evcounts= sort(table(dmg_noaa_table$EVTYPE), decreasing = T)
```

The table contains `numevtypes` different weather events. Ten most frequent events:

```
evcounts[1:10]
```

```
##
##              HAIL        TSTM WIND   THUNDERSTORM WIND
##            288661           219940               82563
##           TORNADO      FLASH FLOOD               FLOOD
##             60652            54277               25326
## THUNDERSTORM WINDS        HIGH WIND           LIGHTNING
##             20843            20212               15754
##        HEAVY SNOW
##             15708
```

Results

```
summary(cars)
```

```
##     speed           dist
##  Min.   : 4.0   Min.   :  2.00
##  1st Qu.:12.0   1st Qu.: 26.00
##  Median :15.0   Median : 36.00
##  Mean   :15.4   Mean   : 42.98
##  3rd Qu.:19.0   3rd Qu.: 56.00
##  Max.   :25.0   Max.   :120.00
```