Impact of Weather Events on the Economy and Population Health in U.S.A

Synopsis

Weather events, such as tornadoes, storms and flood have caused a lot of trouble on public health and economic consequences.

This report will show what type of the weather events have the most impact on population and Economy in U.S.A.

The data was collected from 1950 to Novenber 2011.

Data Processing

Reading Data

- Data is downloaded from: https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2
- The name of data flie is "repdata-data-StormData.csv.bz2".

```
Sys.setlocale('LC_ALL', 'English') #If your system is not in English.
```

```
## [1] "LC_COLLATE=English_United States.1252;LC_CTYPE=English_United States.1252;LC_MONETARY=English_U
```

```
stormdata=read.csv("repdata-data-StormData.csv.bz2")
```

Data Transformations

Time from 1950 to Nov. 2011

```
stormdata$BGN_DATE=as.Date(stormdata$BGN_DATE, format = "%m/%d/%Y")
stormdata.after1950= stormdata[(stormdata$BGN_DATE >= "1950-01-01"), ]
str(stormdata.after1950$BGN_DATE)
```

```
##  Date[1:902297], format: "1950-04-18" "1950-04-18" "1951-02-20" "1951-06-08" ...
```

Thus, we have 902297 observations and 37 variables:

```
dim(stormdata.after1950)
```

```
## [1] 902297     37
```

Loading required packages

```
library(Hmisc)
```

```
## Loading required package: grid
## Loading required package: lattice
## Loading required package: survival
## Loading required package: splines
## Loading required package: Formula
##
## Attaching package: 'Hmisc'
##
## The following objects are masked from 'package:base':
##
##     format.pval, round.POSIXt, trunc.POSIXt, units
```

```
library(reshape)
library(car)
library(lattice)
library(ggplot2)
```

Regulating the events' types(EVTYPE)

Change the EVTYPE into small letter:

```
stormdata$EVTYPE=capitalize(tolower(stormdata$EVTYPE))
```

The Data is then processed and analyzed to: - The effect on *population health*: "pop.health" dataframe which is summrized by fatalities and injuries. - The effect on *economy*: "eco.dam" dataframe which is summrized by properity damage and crop damage. - "pop.health.top" and "eco.dam.top"" will show top 10 events.

The effect on *population health*:

```
pop.health=aggregate(cbind(FATALITIES, INJURIES) ~ EVTYPE, stormdata, sum)
```

Top 10 for each event:

```
ph1=head(pop.health[order(-pop.health$FATALITIES),c(1,2) ], 10)
ph2=head(pop.health[order(-pop.health$INJURIES),c(1,3) ], 10)
ph1
```

```
##             EVTYPE FATALITIES
## 758        Tornado       5633
## 116 Excessive heat       1903
## 138    Flash flood        978
## 243           Heat        937
## 418      Lightning        816
## 779      Tstm wind        504
## 154          Flood        470
## 524    Rip current        368
## 320      High wind        248
## 19       Avalanche        224
```

```
ph2
```

```
##                EVTYPE INJURIES
## 758           Tornado    91346
## 779         Tstm wind     6957
## 154             Flood     6789
## 116    Excessive heat     6525
## 418         Lightning     5230
## 243              Heat     2100
## 387         Ice storm     1975
## 138       Flash flood     1777
## 685 Thunderstorm wind     1488
## 212              Hail     1361
```

Top 10 for the sum:

```
ph.sum=aggregate(FATALITIES+INJURIES ~ EVTYPE, stormdata, sum)
head(ph.sum[order(-ph.sum[,2]),c(1,2) ], 10)
```

```
##                EVTYPE FATALITIES + INJURIES
## 758           Tornado                 96979
## 116    Excessive heat                  8428
## 779         Tstm wind                  7461
## 154             Flood                  7259
## 418         Lightning                  6046
## 243              Heat                  3037
## 138       Flash flood                  2755
## 387         Ice storm                  2064
## 685 Thunderstorm wind                  1621
## 888      Winter storm                  1527
```

The effect on *economy*:

First, we convert PROPDMG*PROPDMGEXP and CROPDMG*CROPDMGEXP into the format of "dollar"

```
stormdata$PROPDMG <- stormdata$PROPDMG * as.numeric(Recode(stormdata$PROPDMGEXP,
    "'0'=1;'1'=10;'2'=100;'3'=1000;'4'=10000;'5'=100000;'6'=1000000;'7'=10000000;'8'=100000000;'B'=1000
    as.factor.result = FALSE))
stormdata$CROPDMG <- stormdata$CROPDMG * as.numeric(Recode(stormdata$CROPDMGEXP,
    "'0'=1;'2'=100;'B'=1000000000;'k'=1000;'K'=1000;'m'=1000000;'M'=1000000;''=0;'?'=0",
    as.factor.result = FALSE))
```

Then, we have effect and top10:

```
eco.dam=aggregate(cbind(PROPDMG, CROPDMG) ~ EVTYPE, stormdata, sum)
```

Top 10 for each event:

```
head(eco.dam[order(-eco.dam$PROPDMG),c(1,2) ], 10)
```

```
##                EVTYPE   PROPDMG
## 59              Flood 1.447e+11
## 167 Hurricane/typhoon 6.931e+10
## 309           Tornado 5.695e+10
## 260       Storm surge 4.332e+10
## 47        Flash flood 1.682e+10
## 93               Hail 1.574e+10
## 159         Hurricane 1.187e+10
## 317    Tropical storm 7.704e+09
## 372      Winter storm 6.688e+09
## 144         High wind 5.270e+09
```

```
head(eco.dam[order(-eco.dam$CROPDMG),c(1,3) ], 10)
```

```
##                EVTYPE   CROPDMG
## 59              Flood 5.171e+09
## 226       River flood 5.029e+09
## 175         Ice storm 5.022e+09
## 159         Hurricane 2.689e+09
## 167 Hurricane/typhoon 2.608e+09
## 93               Hail 2.054e+09
## 32             Drought 1.653e+09
## 47        Flash flood 1.388e+09
## 79       Frost/freeze 9.319e+08
## 144         High wind 6.319e+08
```

Top 10 for the sum:

```
eco.sum=aggregate(PROPDMG+CROPDMG ~ EVTYPE, stormdata, sum)
head(eco.sum[order(-eco.sum[,2]),c(1,2) ], 10)
```

```
##                EVTYPE PROPDMG + CROPDMG
## 59              Flood         1.498e+11
## 167 Hurricane/typhoon         7.191e+10
## 309           Tornado         5.735e+10
## 260       Storm surge         4.332e+10
## 47        Flash flood         1.821e+10
## 93               Hail         1.779e+10
## 159         Hurricane         1.456e+10
## 226       River flood         1.015e+10
## 175         Ice storm         8.967e+09
## 317    Tropical storm         8.156e+09
```

Results

Population Health:

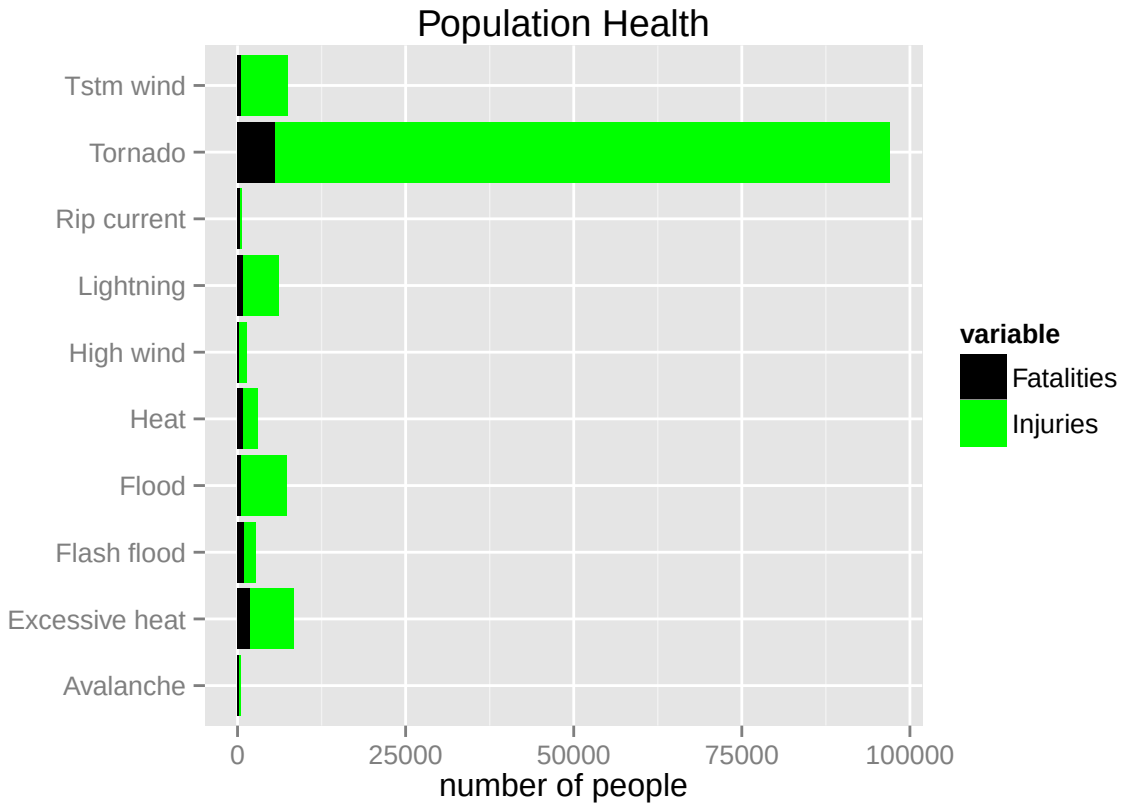
Question:

```
pop.health.top.melt=melt(head(pop.health[order(-pop.health$FATALITIES, -pop.health$INJURIES), ], 10))
```

Across the United States, which types of events (as indicated in the EVTYPE variable) are most harmful with respect to population health?

```
## Using EVTYPE as id variables
```

```
ggplot(pop.health.top.melt, aes(x = EVTYPE, y = value, fill = variable)) + geom_bar(stat = "identity") +
    coord_flip() + ggtitle("Population Health") + labs(x = "", y = "number of people") +
    scale_fill_manual(values = c("black", "green"), labels = c("Fatalities", "Injuries"))
```



Economic Consequences

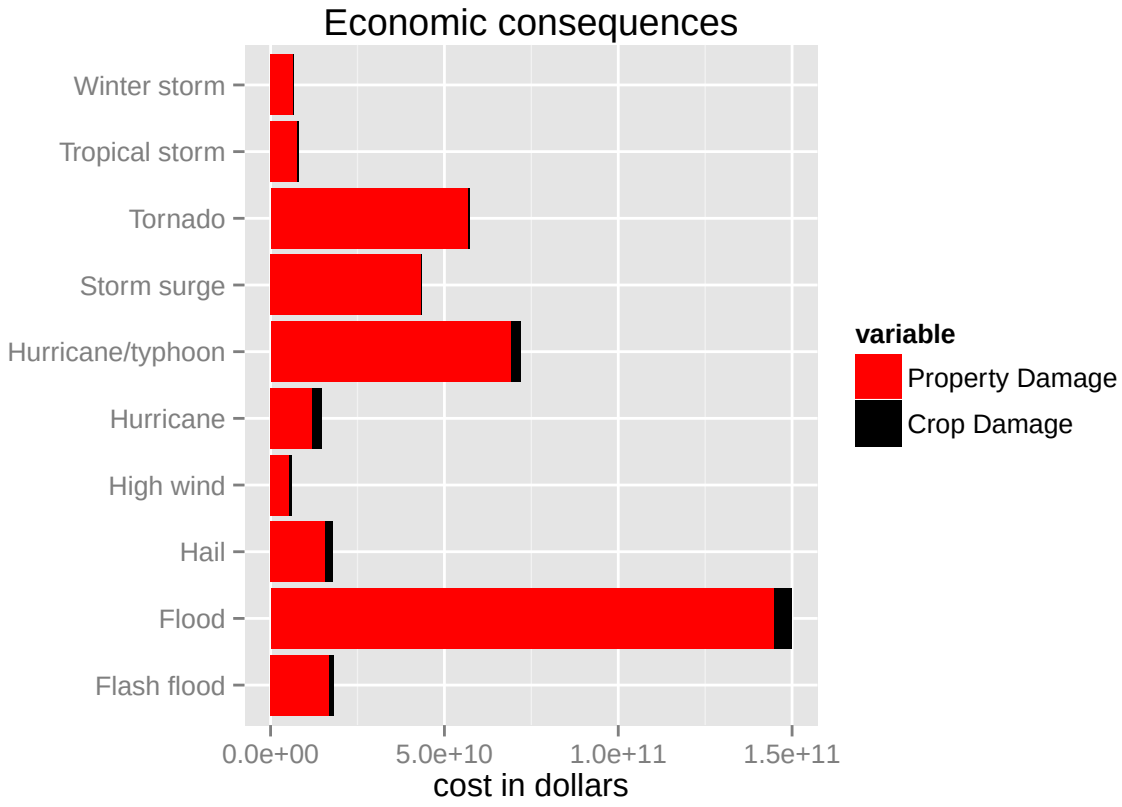
Question:

```
eco.dam.top.melt=melt(head(eco.dam[order(-eco.dam$PROPDMG, -eco.dam$CROPDMG), ], 10))
```

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

```
## Using EVTYPE as id variables
```

```
ggplot(eco.dam.top.melt, aes(x = EVTYPE, y = value, fill = variable)) + geom_bar(stat = "identity") +
    coord_flip() + ggtitle("Economic consequences") + labs(x = "", y = "cost in dollars") +
    scale_fill_manual(values = c("red", "black"), labels = c("Property Damage", "Crop Damage"))
```



From the above plots, we can conclude that:

- The weather event which has the most impact on *population health* is **Tornado**.
- The weather event which has the most impact on *economic consequences* is **Flood**.