# The Most Harmful Types of Weather Events

## Synopsis

In this report I am going to investigate what types of weather events are the most harmful with respect to population health and economic consequences across the United States in our recorded history. This question can be answered in terms of many ways, and I think answering it both in average and in total is a very good start.

## Data Processing

In this assignment I am given the data from the U.S. National Oceanic and Atmospheric Administration's (NOAA) storm database. This database tracks characteristics of major storms and weather events in the United States, including when and where they occur, as well as estimates of any fatalities, injuries, property damage, and crop damage.

First let us read the data in R. The data come in the form of a comma-separated-value file. The events in the database start in the year 1950 and end in November 2011. In the earlier years of the database there are generally fewer events recorded, most likely due to a lack of good record. More recent years should be considered more complete. So it may be better to consider the recent years for our question, but the requirement of this assignment does not say so, and I think I'd better not go into that way to avoid deviation from the requirement.

```
storm<-read.csv(bzfile("repdata_data_StormData.csv.bz2"))
```

The data is ready to use, and we can go directly into data analysis without any other processing. Take a look at the dataset, and you will see that, among all the variables in this dataset, only the variables `FATALITIES`, `INJURIES`,`PROPDMG`, and `CROPDMG` are related to population health and economic consequences, and they indicate respectively the fatalities, injuries, property damage, and crop damage caused by the weather events.

```
summary(storm$FATALITIES)
```

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
##    0.0000    0.0000    0.0000    0.0168    0.0000 583.0000
```

```
summary(storm$INJURIES)
```

```
##       Min.    1st Qu.     Median       Mean    3rd Qu.       Max.
##     0.0000     0.0000     0.0000     0.1557     0.0000 1700.0000
```

```
summary(storm$PROPDMG)
```

```
##     Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##     0.00    0.00    0.00   12.06    0.50 5000.00
```

```
summary(storm$CROPDMG)
```

```
##     Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    0.000   0.000   0.000   1.527   0.000 990.000
```

There is no missing values or abnormal values for those variables. Everything looks good. But you may notice that damage is recorded separately as property damage and crop damage. To measure the economic consequences caused by weather events, I think I need also consider the total damage, thus I sum them up to get this variable.

```
storm$DMG<-storm$PROPDMG+storm$CROPDMG
```

## Results

First let's look at the most harmful weather events in terms of average population health and economic consequences. Thus I calculate those averages of fatalities, injuries, property damages, crop damages, and total damages by event type.

```
death<-tapply(storm$FATALITIES,storm$EVTYPE,mean)
injury<-tapply(storm$INJURIES,storm$EVTYPE,mean)
propdmg<-tapply(storm$PROPDMG,storm$EVTYPE,mean)
cropdmg<-tapply(storm$CROPDMG,storm$EVTYPE,mean)
dmg<-tapply(storm$DMG,storm$EVTYPE,mean)
```

The most harmful weather events with respect to fatalities are:

```
head(sort(death,decreasing=T))
```

```
## TORNADOES, TSTM WIND, HAIL              COLD AND SNOW
##                  25.000000                  14.000000
##      TROPICAL STORM GORDON      RECORD/EXCESSIVE HEAT
##                   8.000000                   5.666667
##               EXTREME HEAT          HEAT WAVE DROUGHT
##                   4.363636                   4.000000
```

Although we see `TORNADOES, THUNDERSTORM WIND, HAIL` is the most harmful weather event causing fatalities in average, one may notice that there are other types of events representing tornado, thunderstorm wind, and hail in the dataset, and it seems that there have been changes in how to record events. One may think of cleaning up the variable `storm$EVTYPE` before investigating the most harmful events. However, we do not need do it. I will explain it later.

Next we'll take a look at injuries.

```
head(sort(injury,decreasing=T))
```

```
##             Heat Wave TROPICAL STORM GORDON            WILD FIRES
##                  70.0                  43.0                  37.5
##         THUNDERSTORMW    HIGH WIND AND SEAS       SNOW/HIGH WINDS
##                  27.0                  20.0                  18.0
```

So the most harmful weather event causing injuries are `Heat Wave`, `TROPICAL STORM GORDON`, `WILD FIRES`, and so on. But I am a little curious about how many injuries were caused by `TORNADOES, THUNDERSTORM WIND, HAIL`, which incurred the most fatalities in average in the history.

```
injury[names(injury)=="TORNADOES, TSTM WIND, HAIL"]
```

```
## TORNADOES, TSTM WIND, HAIL
##                          0
```

There is no injury occurred for the most harmful event in average for death. It may be surprising, or not surprising at all.

Next we will look at damages.

```
head(sort(propdmg,decreasing=T))
```

```
##        COASTAL EROSION   HEAVY RAIN AND FLOOD RIVER AND STREAM FLOOD
##                    766                    600                    600
##              Landslump  BLIZZARD/WINTER STORM           FLASH FLOOD/
##                    570                    500                    500
```

```
head(sort(cropdmg,decreasing=T))
```

```
## DUST STORM/HIGH WINDS          FOREST FIRES TROPICAL STORM GORDON
##               500.000               500.000               500.000
##       HIGH WINDS/COLD       HURRICANE FELIX        River Flooding
##               401.000               250.000               241.368
```

```
head(sort(dmg,decreasing=T))
```

```
##  TROPICAL STORM GORDON        COASTAL EROSION   HEAVY RAIN AND FLOOD
##                   1000                    766                    600
## RIVER AND STREAM FLOOD              Landslump  DUST STORM/HIGH WINDS
##                    600                    570                    550
```

The weather events causing the most property damages in average are `COASTAL EROSION`, `HEAVY RAIN AND FLOOD`, `RIVER AND STREAM FLOOD`, and so on, and those incurring the most crop damages in average are `DUST STORM/HIGH WINDS`, `FOREST FIRES`, `TROPICAL STORM GORDON` and etc., and in total damages, the most harmful events in average are `TROPICAL STORM GORDON`, `COASTAL EROSION`, and so on.

Similarly we can investigate the most harmful weather events in terms of total population health and economic consequences. First get the totals of fatalities, injuries, property damages, crop damages, and total damages by event type.

```
death<-tapply(storm$FATALITIES,storm$EVTYPE,sum)
injury<-tapply(storm$INJURIES,storm$EVTYPE,sum)
propdmg<-tapply(storm$PROPDMG,storm$EVTYPE,sum)
cropdmg<-tapply(storm$CROPDMG,storm$EVTYPE,sum)
dmg<-tapply(storm$DMG,storm$EVTYPE,sum)
```

Then we look at what the most harmful events are in terms of the total fatalities and injuries caused in history.

```
head(sort(death,decreasing=T),10)
```

```
##        TORNADO EXCESSIVE HEAT    FLASH FLOOD           HEAT      LIGHTNING
##           5633           1903            978            937            816
##      TSTM WIND          FLOOD    RIP CURRENT      HIGH WIND      AVALANCHE
##            504            470            368            248            224
```

```
head(sort(injury,decreasing=T),10)
```

```
##           TORNADO         TSTM WIND             FLOOD    EXCESSIVE HEAT
##             91346              6957              6789              6525
##         LIGHTNING              HEAT         ICE STORM       FLASH FLOOD
##              5230              2100              1975              1777
## THUNDERSTORM WIND              HAIL
##              1488              1361
```

Now we can see that tornado is the event causing the most fatalities and injuries in total in the United States in history. You may also notice that the total fatalities and injuries caused by tornado are much more than the second leaders in either aspect.

Then look at the damages.

```
head(sort(propdmg,decreasing=T),10)
```

```
##            TORNADO        FLASH FLOOD          TSTM WIND
##          3212258.2          1420124.6          1335965.6
##              FLOOD  THUNDERSTORM WIND               HAIL
##           899938.5           876844.2           688693.4
##          LIGHTNING THUNDERSTORM WINDS          HIGH WIND
##           603351.8           446293.2           324731.6
##       WINTER STORM
##           132720.6
```

```
head(sort(cropdmg,decreasing=T),10)
```

```
##               HAIL        FLASH FLOOD              FLOOD
##          579596.28          179200.46          168037.88
##          TSTM WIND            TORNADO  THUNDERSTORM WIND
##          109202.60          100018.52           66791.45
##            DROUGHT THUNDERSTORM WINDS          HIGH WIND
##           33898.62           18684.93           17283.21
##         HEAVY RAIN
##           11122.80
```

```
head(sort(dmg,decreasing=T),10)
```

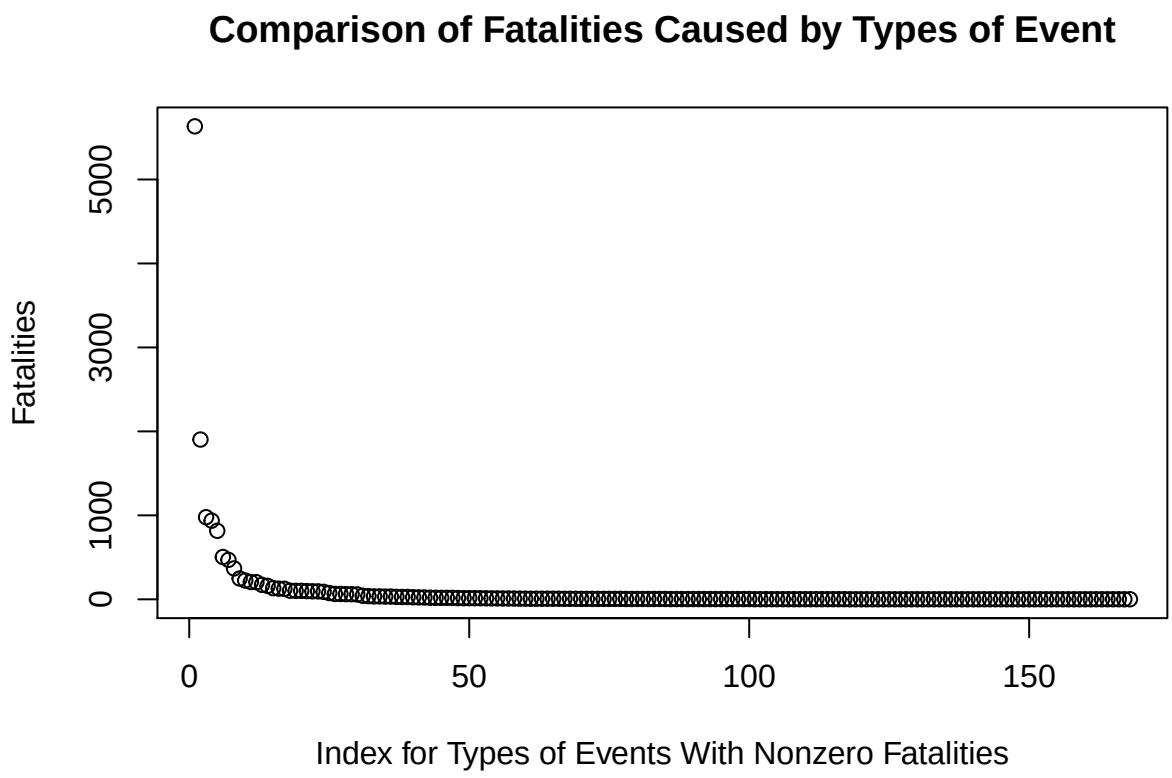
```
##            TORNADO        FLASH FLOOD          TSTM WIND
##          3312276.7          1599325.1          1445168.2
##               HAIL              FLOOD  THUNDERSTORM WIND
##          1268289.7          1067976.4           943635.6
```

```
##          LIGHTNING THUNDERSTORM WINDS          HIGH WIND
##           606932.4           464978.1           342014.8
##       WINTER STORM
##           134699.6
```

Again tornado is the event causing the most property damages and total damages (including property damage and crop damage) in total in the United States in history, while for crop damage, hail takes the lead and tornado becomes the fifth. Other weather events listed in the top (i.e., those causing most damages) are flash flood, thunderstorm wind, and flood.

Again, you may notice that thunderstorm is recorded in at least two different types in the dataset, `TSTM WIND` and `THUNDERSTORM WIND`, a sign that there have probably been changes in how to record events. But we do not need clean up the variable `storm$EVTYPE` before doing data analysis. The reason is that those most harmful weather events in total in population health and economic consequences constitute the most damages to total population health and economic consequences, compared to other weather events recorded. You may understand it very well from the following plots in fatalities and total damages. Here the numbers in the x axis are just some indices created temporarily so you know roughly how many event types there are. Actually you can ignore those indices, because even if a harmful event is recorded multiple times, it will still show up in the top when we ignore the multiple entries, if the harmful event has a lot of damage. The two figures also show that, in order to avoid damages in total population health and economic consequences, we should gear our resources to those most harmful events discovered, which are reported clearly in this report, especially those in terms of total population health and economic consequences (because even each occurrence does not have many damages, but if it occurs often, it is still a concern).

```
plot(sort(death[death>0],decreasing=T),xlab="Index for Types of Events With Nonzero Fatalities",
     ylab="Fatalities",main="Comparison of Fatalities Caused by Types of Event")
```



```
plot(sort(dmg[dmg>0],decreasing=T),xlab="Index for Types of Events With Nonzero Total Damages",
     ylab="Total Damages",main="Comparison of Total Damages Caused by Types of Event")
```