The Impact of Severe Weather Events on the Public Health and Economy

Li Xu

Synopsis

In this report, we use the U.S. National Oceanic and Atmospheric Administration's storm database, and study the impact of severe weather events on population health and economy. We conclude that among all weather events, the tornado is the most harmful event with respect to population health, and the flood has the greatest economic consequences.

Data Processing

In this section, we will download and read the data set, then make data transformation for subsequent data analysis.

Downloading and Reading data

The data set can be downloaded from the following website

```
https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2
```

and we save it in the working directory. We read the data set by typing

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

This data set contains 902297 observations with the following 37 variables.

```
 [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"
```

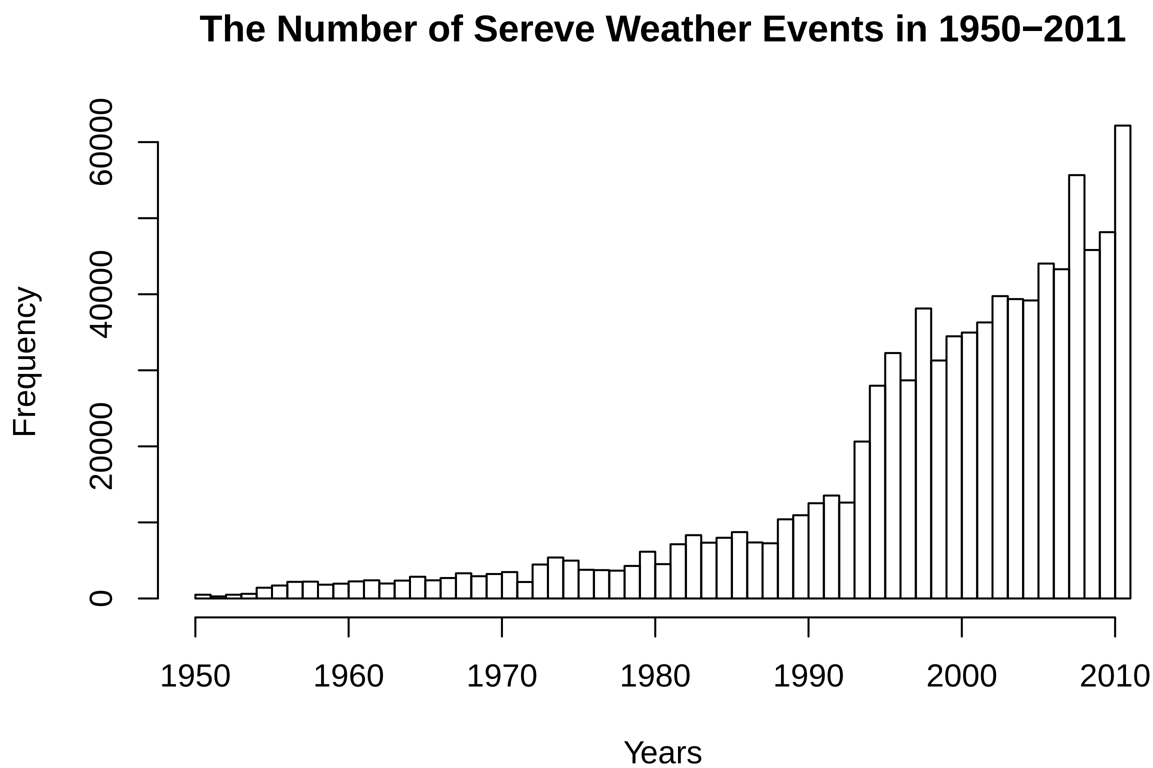
Data Transformation

We need to preprocess the data set "StormData" before we analyze it. First, notice that there are two features about date. We convert the string value of "date" into date format for convenience by typing

```
StormData$BGN_DATE<-as.Date(StormData$BGN_DATE,"%m/%d/%Y")
StormData$END_DATE<-as.Date(StormData$END_DATE,"%m/%d/%Y")
```

We count the number of events in each year and draw a histogram about the frequency of events as follows:

```r
StormData<-read.csv("repdata-data-StormData.csv.bz2")
StormData$BGN_DATE<-as.Date(StormData$BGN_DATE,"%m/%d/%Y")
StormData$END_DATE<-as.Date(StormData$END_DATE,"%m/%d/%Y")
StormData$YEAR<-as.numeric(format(StormData$BGN_DATE,"%Y"))
hist(StormData$YEAR,breaks=60,xlab="Years",ylab="Frequency",
     main="The Number of Sereve Weather Events in 1950-2011")
```



Notice that in the earlier years (1950-1993) of the database, there are fewer events (less than 20000 events per year). So we subset the data set by

```r
ProcessedData<-subset(StormData,StormData$YEAR>=1993)
```

The processed data contains 714738 samples.

Next, We extract four variables "CROPDMG","CROPDMGEXP", "PROPDMG", "PROPDMGEXP" related to economy. Notice that the values of damage are expressed in scientific notation. Due to analysis convenience, we need to convert them into the ordinary form. By typing

```r
levels(ProcessedData$CROPDMGEXP)
levels(ProcessedData$PROPDMGEXP)
```

we see that variable "CROPDMGEXP" contains the following characters

```
[1] ""  "?" "0" "2" "B" "k" "K" "m" "M"
```

and "PROPDMGEXP" contains the following characters

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

Then we convert each character into a power of 10, multiply it by "CROPDMG" or "PROPDMG", and save it in a new column "CROPDMG.merge" or "PROPDMG.merge" The R code is as follows:

```
power<-function(x){
  powerlist<-list("0"=1, "1"=10, "2"=100, "3"=1000, "4"=10000, "5"=100000,
  "6"=1000000, "7"=10000000, "8"=100000000, "B"=1000000000, "h"=100,
  "H"=100, "k"=1000, "K"=1000, "m"=1000000, "M"=1000000, "b"=1000000000,
  "B"=1000000000)
  if(x %in% names(powerlist)){returnvalue<-powerlist[[x]]}
  else{returnvalue<-0}
  return(returnvalue)
}
ProcessedData$CROPDMG.merge<-apply(ProcessedData,1,
    FUN=function(x) as.numeric(x["CROPDMG"])*as.numeric(power(x["CROPDMGEXP"])))
ProcessedData$PROPDMG.merge<-apply(ProcessedData,1,
    FUN=function(x) as.numeric(x["PROPDMG"])*as.numeric(power(x["PROPDMGEXP"])))
```

Results

In the following, we will analyze the data set and answer which sereve weather has the most significant influence on population health and economy.

Most Harmful Events for Population Health

In this section, we study the impact of sereve weather events on the variables "INJURIES" and "FATALITIES", which indicate the number of injuries and fatalities, respectively. First, we subset the data set related to "EVTYPE", "INJURIES" and "FATALITIES" by typing

```
HealthData<-ProcessedData[,c("EVTYPE","INJURIES","FATALITIES")]
```

We aggregate the number of injuries and fatalities for each event type via typing

```
AgHealthData<-aggregate(.~EVTYPE,data=HealthData,FUN=sum,na.rm=T)
```

Then via

```
OrderedAgHealthData.Inj<-AgHealthData[order(AgHealthData$INJURIES,decreasing=T),]
OrderedAgHealthData.Inj[1:10,]
```

we order the number of injuries by weather events and pick the top ten events causing the most injuries as follows:

```
               EVTYPE INJURIES FATALITIES
834           TORNADO    23310       1621
170             FLOOD     6789        470
130    EXCESSIVE HEAT     6525       1903
464         LIGHTNING     5230        816
```

```
856        TSTM WIND     3631        241
275             HEAT     2100        937
427        ICE STORM     1975         89
153      FLASH FLOOD     1777        978
760 THUNDERSTORM WIND    1488        133
972     WINTER STORM     1321        206
```

Similarly, by typing

```
OrderedAgHealthData.Fatal<-AgHealthData[order(AgHealthData$FATALITIES,decreasing=T),]
OrderedAgHealthData.Fatal[1:10,]
```

we can also list the top ten events causing the most fatalities

```
            EVTYPE INJURIES FATALITIES
130 EXCESSIVE HEAT     6525       1903
834        TORNADO    23310       1621
153    FLASH FLOOD     1777        978
275           HEAT     2100        937
464      LIGHTNING     5230        816
170          FLOOD     6789        470
585    RIP CURRENT      232        368
359      HIGH WIND     1137        248
856      TSTM WIND     3631        241
19       AVALANCHE      170        224
```

Finally, we combine the numbers of injuries and fatalties, and then list the top ten harmful events leading to the most sum of injuries and fatalties. The R code is

```
AgHealthData$SUM<-AgHealthData$INJURIES+AgHealthData$FATALITIES
OrderedAgHealthData.Both<-AgHealthData[order(AgHealthData$SUM,decreasing=T),]
OrderedAgHealthData.Both[1:10,]
```

We then obtain the following result

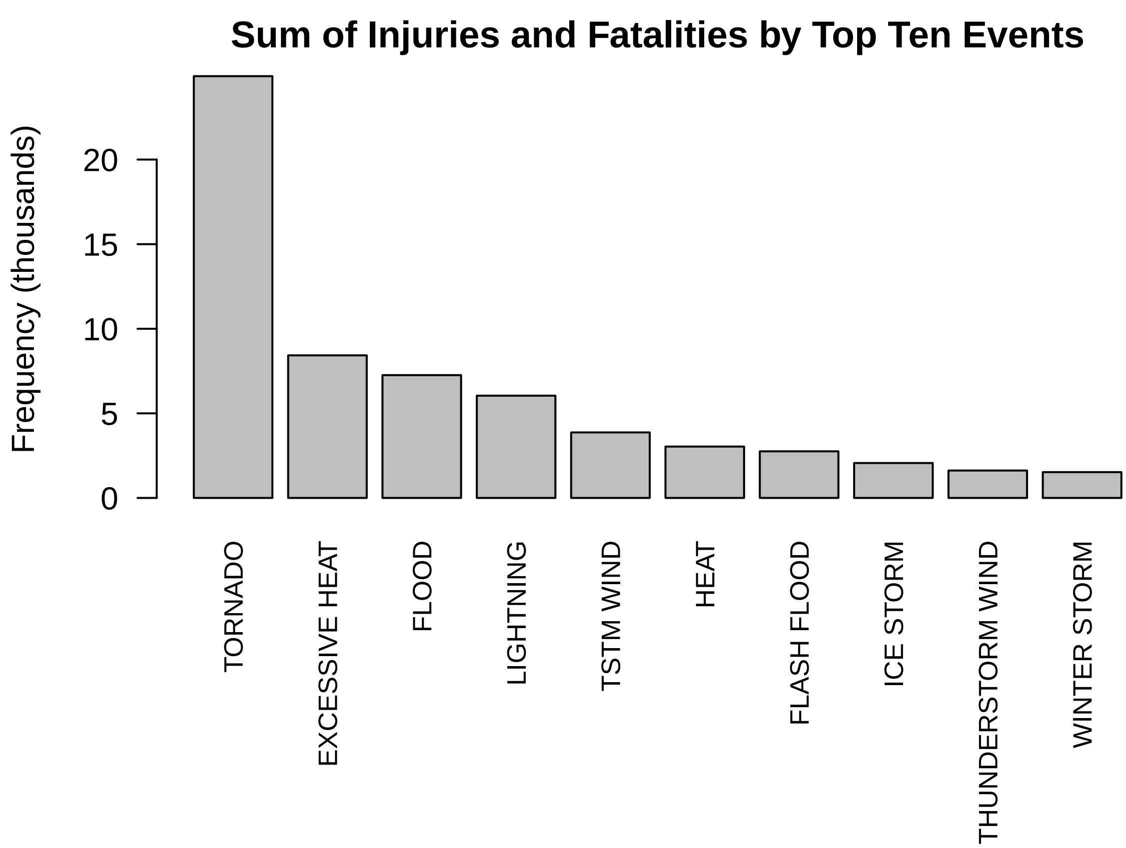
```
               EVTYPE INJURIES FATALITIES   SUM
834           TORNADO    23310       1621 24931
130    EXCESSIVE HEAT     6525       1903  8428
170             FLOOD     6789        470  7259
464         LIGHTNING     5230        816  6046
856         TSTM WIND     3631        241  3872
275              HEAT     2100        937  3037
153       FLASH FLOOD     1777        978  2755
427         ICE STORM     1975         89  2064
760 THUNDERSTORM WIND     1488        133  1621
972      WINTER STORM     1321        206  1527
```

By typing

```
par(mar = c(9, 4, 2, 2) + 0.2)
barplot(OrderedAgHealthData.Both[1:10,]$SUM/1000, horiz=F, cex.names=0.8, xlab="",
    ylab="Frequency (thousands)",las=2,names.arg=OrderedAgHealthData.Both[1:10,]$EVTYPE,
    main="Sum of Injuries and Fatalitie0s by Top Ten Events")
```

We make a bar plot for the top ten events

```
setwd("C:/Users/xuli/Desktop/specialization/5/p2")
StormData<-read.csv("repdata-data-StormData.csv.bz2")
StormData$BGN_DATE<-as.Date(StormData$BGN_DATE,"%m/%d/%Y")
StormData$END_DATE<-as.Date(StormData$END_DATE,"%m/%d/%Y")
StormData$YEAR<-as.numeric(format(StormData$BGN_DATE,"%Y"))
ProcessedData<-subset(StormData,StormData$YEAR>=1993)
HealthData<-ProcessedData[,c("EVTYPE","INJURIES","FATALITIES")]
AgHealthData<-aggregate(.~EVTYPE,data=HealthData,FUN=sum,na.rm=T)
AgHealthData$SUM<-AgHealthData$INJURIES+AgHealthData$FATALITIES
OrderedAgHealthData.Both<-AgHealthData[order(AgHealthData$SUM,
                                             decreasing=T),]
par(mar = c(9, 4, 2, 2) + 0.2)
barplot(OrderedAgHealthData.Both[1:10,]$SUM/1000, horiz=F, cex.names=0.8, xlab="",
    ylab="Frequency (thousands)",las=2,names.arg=OrderedAgHealthData.Both[1:10,]$EVTYPE,
    main="Sum of Injuries and Fatalities by Top Ten Events")
```



Hence, we can conclude that **tornado is the most harmful event with respect to population health**.

Economic Consequences of Sereve Weather Events

Next, we study the impact of sereve Weather Events on the economy. We extract two merged variables "CROPDMG.merge", "PROPDMG.merge" related to economy by

```r
EconomyData<-ProcessedData[,c("EVTYPE","CROPDMG.merge","PROPDMG.merge")]
```

We aggregate the amount of crop damage and property damage for each event type via typing

```r
AgEconomyData<-aggregate(.~EVTYPE,data=EconomyData,FUN=sum,na.rm=T)
```

Then with the following R code

```r
OrderedAgEconomyData.Crop<-AgEconomyData[order(AgEconomyData$CROPDMG.merge,decreasing=T),]
OrderedAgEconomyData.Crop[1:10,]
```

we can order the crop damage by events and pick the top ten events causing the most crop damage as follows:

```
               EVTYPE CROPDMG.merge PROPDMG.merge
95            DROUGHT   13972566000    1046106000
170             FLOOD    5661968450  144657709800
590       RIVER FLOOD    5029459000    5118945500
427         ICE STORM    5022113500    3944927860
244              HAIL    3025954470   15735267456
402         HURRICANE    2741910000   11868319010
411 HURRICANE/TYPHOON    2607872800   69305840000
153       FLASH FLOOD    1421317100   16822673772
140      EXTREME COLD    1292973000      67737400
212      FROST/FREEZE    1094086000       9480000
```

Similarly, by typing

```r
OrderedAgEconomyData.Prop<-AgEconomyData[order(AgEconomyData$PROPDMG.merge,decreasing=T),]
OrderedAgEconomyData.Prop[1:10,]
```

we can also list top ten events causing the most property damage

```
               EVTYPE CROPDMG.merge PROPDMG.merge
170             FLOOD    5661968450  144657709800
411 HURRICANE/TYPHOON    2607872800   69305840000
670       STORM SURGE          5000   43323536000
834           TORNADO     414953270   26349182044
153       FLASH FLOOD    1421317100   16822673772
244              HAIL    3025954470   15735267456
402         HURRICANE    2741910000   11868319010
848    TROPICAL STORM     678346000    7703890550
972      WINTER STORM      26944000    6688497251
359         HIGH WIND     638571300    5270046260
```

Finally, we sum the amount of crop damage and property damage, and list the top ten harmful events leading to the most sum of two kinds of damage.

```r
AgEconomyData$BOTH<-AgEconomyData$CROPDMG.merge+AgEconomyData$PROPDMG.merge
OrderedAgEconomyData.Both<-AgEconomyData[order(AgEconomyData$BOTH,decreasing=T),]
OrderedAgEconomyData.Both[1:10,]
```

Then we obtain the following result

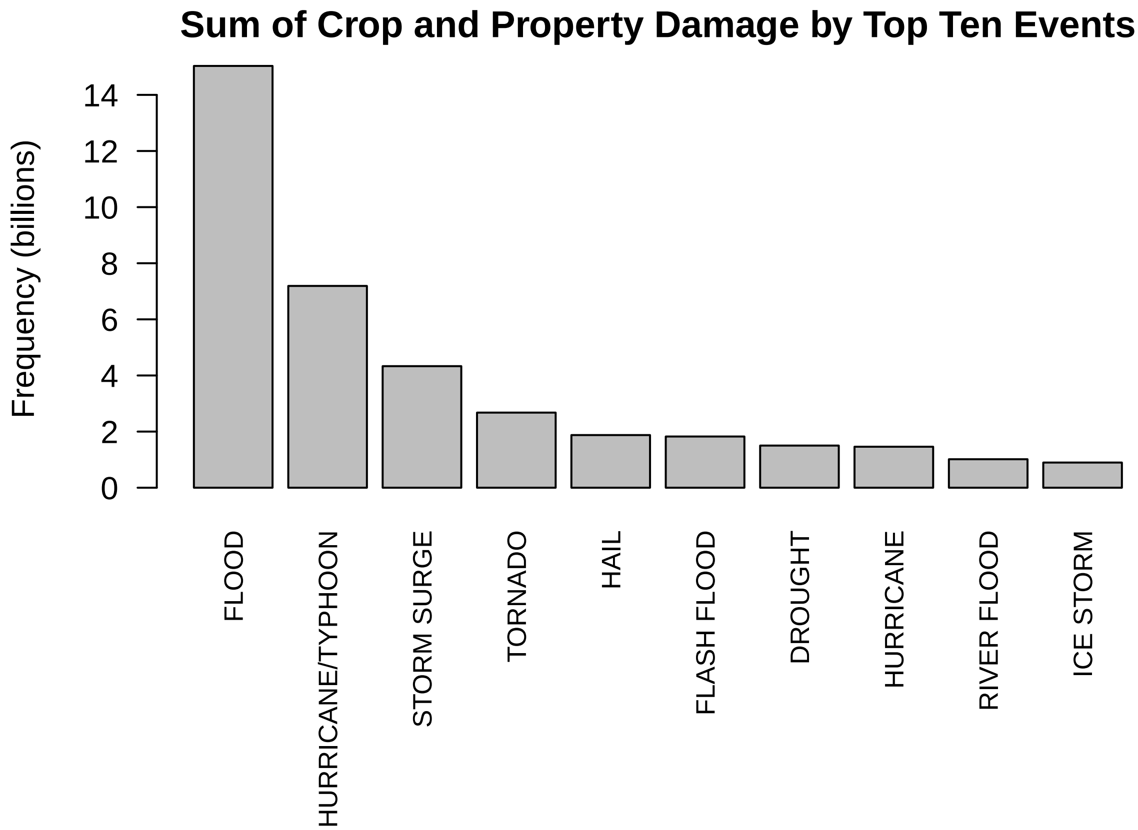
```
                 EVTYPE CROPDMG.merge PROPDMG.merge         BOTH
170               FLOOD    5661968450  144657709800 150319678250
411   HURRICANE/TYPHOON    2607872800   69305840000  71913712800
670         STORM SURGE          5000   43323536000  43323541000
834             TORNADO     414953270   26349182044  26764135314
244                HAIL    3025954470   15735267456  18761221926
153         FLASH FLOOD    1421317100   16822673772  18243990872
95              DROUGHT   13972566000    1046106000  15018672000
402           HURRICANE    2741910000   11868319010  14610229010
590         RIVER FLOOD    5029459000    5118945500  10148404500
427           ICE STORM    5022113500    3944927860   8967041360
```

By typing

```r
par(mar = c(9, 4, 2, 2) + 0.2)
barplot(OrderedAgEconomyData.Both[1:10,]$BOTH/10e9, horiz=F, cex.names=0.8, xlab="",
    ylab="Frequency (billions)",las=2,names.arg=OrderedAgEconomyData.Both[1:10,]$EVTYPE,
    main="Sum of Crop and Property Damage by Top Ten Events")
```

We make a bar plot for the top ten weather events

```r
setwd("C:/Users/xuli/Desktop/specialization/5/p2")
StormData<-read.csv("repdata-data-StormData.csv.bz2")
StormData$BGN_DATE<-as.Date(StormData$BGN_DATE,"%m/%d/%Y")
StormData$END_DATE<-as.Date(StormData$END_DATE,"%m/%d/%Y")
StormData$YEAR<-as.numeric(format(StormData$BGN_DATE,"%Y"))
ProcessedData<-subset(StormData,StormData$YEAR>=1993)
power<-function(x){
  powerlist<-list("0"=1, "1"=10, "2"=100, "3"=1000, "4"=10000, "5"=100000,
  "6"=1000000, "7"=10000000, "8"=100000000, "B"=1000000000, "h"=100,
  "H"=100, "k"=1000, "K"=1000, "m"=1000000, "M"=1000000, "b"=1000000000,
  "B"=1000000000)
  if(x %in% names(powerlist)){returnvalue<-powerlist[[x]]}
  else{returnvalue<-0}
  return(returnvalue)
}
ProcessedData$CROPDMG.merge<-apply(ProcessedData,1,
  FUN=function(x) as.numeric(x["CROPDMG"])*as.numeric(power(x["CROPDMGEXP"])))
ProcessedData$PROPDMG.merge<-apply(ProcessedData,1,
  FUN=function(x) as.numeric(x["PROPDMG"])*as.numeric(power(x["PROPDMGEXP"])))
EconomyData<-ProcessedData[,c("EVTYPE","CROPDMG.merge","PROPDMG.merge")]
AgEconomyData<-aggregate(.~EVTYPE,data=EconomyData,FUN=sum,na.rm=T)
AgEconomyData$BOTH<-AgEconomyData$CROPDMG.merge+AgEconomyData$PROPDMG.merge
OrderedAgEconomyData.Both<-AgEconomyData[order(AgEconomyData$BOTH,decreasing=T),]
par(mar = c(9, 4, 2, 2) + 0.2)
barplot(OrderedAgEconomyData.Both[1:10,]$BOTH/10e9, horiz=F, cex.names=0.8, xlab="",
    ylab="Frequency (billions)",las=2,names.arg=OrderedAgEconomyData.Both[1:10,]$EVTYPE,
    main="Sum of Crop and Property Damage by Top Ten Events")
```



So we can conclude that **flood has the greatest economic consequences**.