An analysis of the economic and human cost of weather events from 1950 onward

Synopsis

In this analysis we examine a data set documenting storm weather events since 1950. The events include several different phenomena such as tornadoes, hail, ect. The data set documents many details for each event such as location and duration. In this analysis we load the dataset into R, remove any non complete cases and extract the fatalities, injuries and economic damage with respect to each event. We then find the the top 10 events the fatalities, injuries and economic damage, and present these findings in tables and bar charts.

Data Processing

First we read the table into R:

```
strmdata<-read.csv("repdata_data_StormData.csv.bz2")
names(strmdata)
```

```
##  [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"
```

Upon examining the data variables we can see that many of the variables are superfluous to what we need so we will remove all variables except “EVTPE”, “FATALITIES”, “INJURIES”, “PROPDMD”. “PROPDMGEXP”, “CROPDMG” and “”CROPDMG“”. This is seen in the following code, we also remove any incomplete cases.

```
workdata<-strmdata[,c("EVTYPE","FATALITIES","INJURIES" ,"PROPDMG","PROPDMGEXP","CROPDM
G","CROPDMGEXP")]
workdata<-na.omit(workdata)
```

The cost of damages is also attributed to a given event, each value is quantified in the “PROPDMGEXP” column, to further process the data we will instead use this column and the “PROPDMG” and create a new column with the numerical values of cost of the damage. If we apply table to “PROPDMGEXP” we see there are other values than “K” (thousands), “M” (million), and “B” billions. Compared to “K,”M" , and “B” we see the other values are not very frequent and they are not explained in the other documents, thus we will just set these values to 1 (so that the value in “PROPDMG” represents just that value, as opposed to a multiple of a thousand ect.). We also convert “K”, “M” and “B” to their numerical equivalents such that “PROPDMG” can be multiplied by these values to get the actual value of damage. This is achieved by the following code:

```
table(workdata$"PROPDMGEXP")
```

```
##
##             -      ?      +      0      1      2      3      4      5
## 465934      1      8      5    216     25     13      4      4     28
##      6      7      8      B      h      H      K      m      M
##      4      5      1     40      1      6 424665      7  11330
```

```
temp<-as.character(workdata$"PROPDMGEXP")
temp[!(temp=="K"|temp=="M"|temp=="B")]<-1
temp[temp=="K"]<-1000
temp[temp=="M"]<-1000000
temp[temp=="B"]<-1000000000
temp<-as.numeric(temp)
ACTCOST<-workdata$"PROPDMG"*temp
workdata<-mutate(workdata,PROPDMGNUM=ACTCOST)
# Now do the same for crop damage
temp<-as.character(workdata$"CROPDMGEXP")
temp[!(temp=="K"|temp=="M"|temp=="B")]<-1
temp[temp=="K"]<-1000
temp[temp=="M"]<-1000000
temp[temp=="B"]<-1000000000
temp<-as.numeric(temp)
ACTCOST<-workdata$"CROPDMG"*temp
workdata<-mutate(workdata,CROPDMGNUM=ACTCOST)
```

We can now use workdata as our data set to determine the cost of various storm disasters

Results

Storm Events and Public Health

When presenting the data we will consider a fatality worse than an injury regardless of quantity (i.e even one fatality is worse than many injuries. To find the total fatalities and injuries for a given disaster we use the following code:

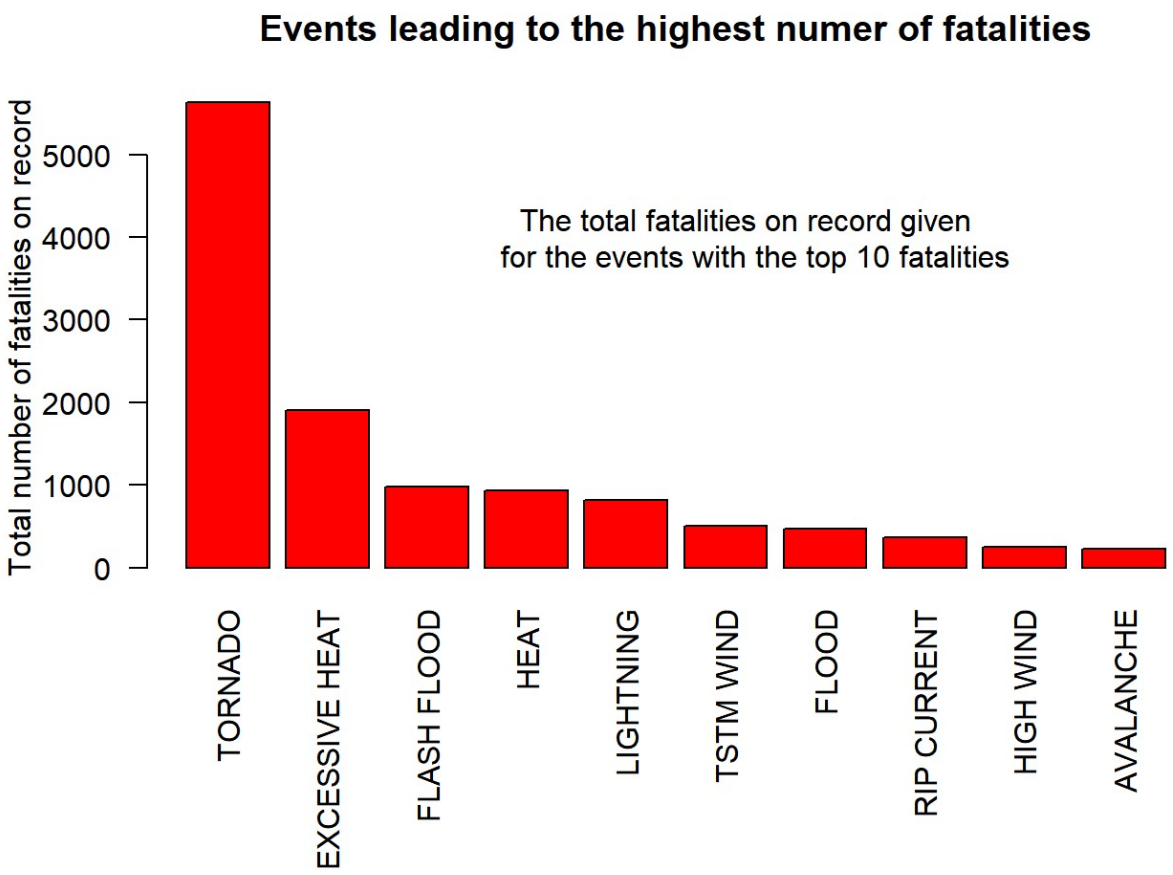
```
fat_inj<-workdata[,c("EVTYPE","FATALITIES","INJURIES")]
event_group<-group_by(fat_inj,EVTYPE)
sums<-summarize_all(event_group,sum)
sums<-arrange(sums,desc(FATALITIES))
sums1<-arrange(sums,desc(INJURIES))
par(mar=c(8, 4, 4, 2) + 0.1)
print(sums)
```

```
## # A tibble: 985 x 3
##    EVTYPE         FATALITIES INJURIES
##    <fct>               <dbl>    <dbl>
##  1 TORNADO              5633    91346
##  2 EXCESSIVE HEAT       1903     6525
##  3 FLASH FLOOD           978     1777
##  4 HEAT                  937     2100
##  5 LIGHTNING             816     5230
##  6 TSTM WIND             504     6957
##  7 FLOOD                 470     6789
##  8 RIP CURRENT           368      232
##  9 HIGH WIND             248     1137
## 10 AVALANCHE             224      170
## # ... with 975 more rows
```

```
print(sums1)
```

```
## # A tibble: 985 x 3
##    EVTYPE            FATALITIES INJURIES
##    <fct>                  <dbl>    <dbl>
##  1 TORNADO                 5633    91346
##  2 TSTM WIND                504     6957
##  3 FLOOD                    470     6789
##  4 EXCESSIVE HEAT          1903     6525
##  5 LIGHTNING                816     5230
##  6 HEAT                     937     2100
##  7 ICE STORM                 89     1975
##  8 FLASH FLOOD              978     1777
##  9 THUNDERSTORM WIND        133     1488
## 10 HAIL                      15     1361
## # ... with 975 more rows
```

```
barplot(sums[[2]][1:10],names.arg =sums[[1]][1:10],las=2, col="red",ylab="Total number
of fatalities on record", main = "Events leading to the highest numer of fatalities")
text(x=7, y=4000,labels = "The total fatalities on record given \n for the events wit
h the top 10 fatalities" )
```



There are almost a thousand events so we show the the top 10 events for both total fatalities and injuries in descending order, we also show a bar graph for the number of fatalities. We see that both lists are quite similar (which makes sense as fatalities and injuries are probably correlated). We also see that tornadoes are a significantly higher source of fatalities and injuries compared to other events.

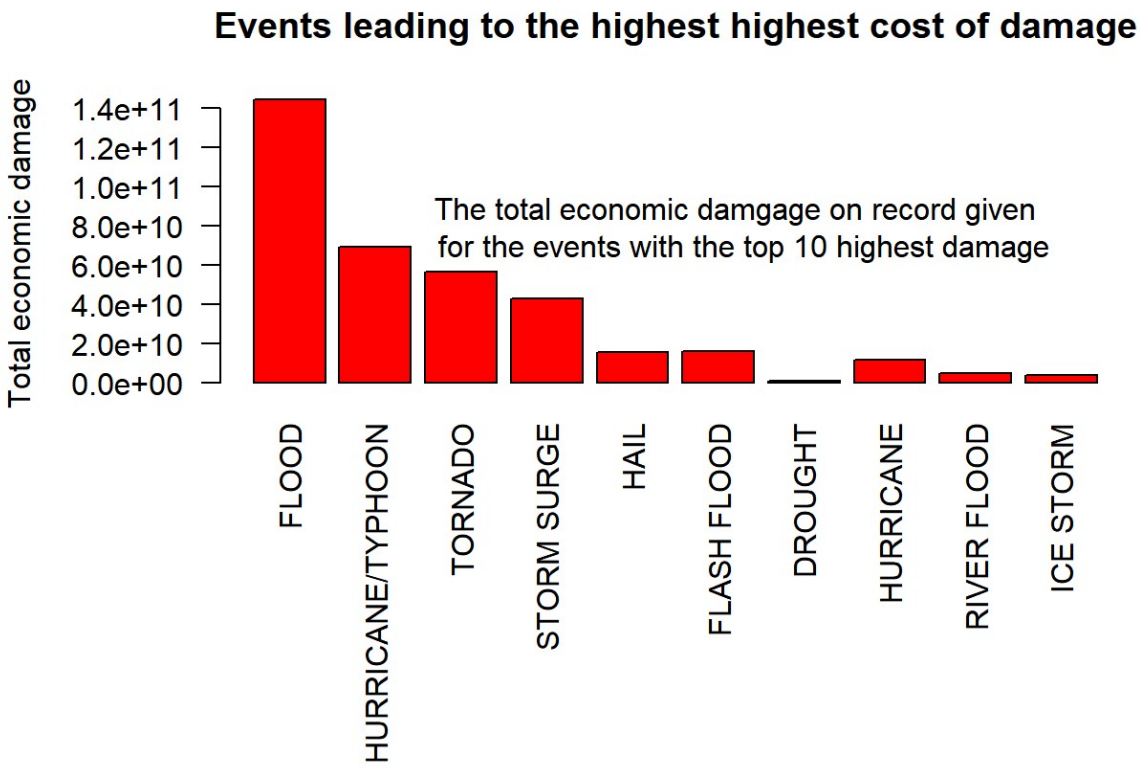
Storm Events and Economic Damage

Now we will look at the events that cause the most economic damage, the economic damage is calculated as the sum of the property damage and the crop damage.

```
dmg<-workdata[,c("EVTYPE","PROPDMGNUM","CROPDMGNUM")]
dmg<-mutate(dmg,TOTALDMG= PROPDMGNUM + CROPDMGNUM )
event_group<-group_by(dmg,EVTYPE)
sums<-summarize_all(event_group,sum)
sums<-arrange(sums,desc(TOTALDMG))
par(mar=c(13, 8, 4, 2) + 0.1)
print(sums)
```

```
## # A tibble: 985 x 4
##    EVTYPE               PROPDMGNUM  CROPDMGNUM      TOTALDMG
##    <fct>                     <dbl>       <dbl>         <dbl>
##  1 FLOOD             144657709807   5661968450 150319678257
##  2 HURRICANE/TYPHOON  69305840000   2607872800  71913712800
##  3 TORNADO            56925660790.   414953270  57340614060.
##  4 STORM SURGE        43323536000         5000  43323541000
##  5 HAIL               15727367053.  3025537890  18752904943.
##  6 FLASH FLOOD        16140812067.  1421317100  17562129167.
##  7 DROUGHT             1046106000  13972566000  15018672000
##  8 HURRICANE          11868319010   2741910000  14610229010
##  9 RIVER FLOOD         5118945500   5029459000  10148404500
## 10 ICE STORM           3944927860   5022113500   8967041360
## # ... with 975 more rows
```

```
barplot(sums[[2]][1:10],names.arg =sums[[1]][1:10],las=2, col="red",ylab="Total econom
ic damage", main = "Events leading to the highest highest cost of damage",mgp = c(5,1,
0))
text(x=7, y=8.0e10,labels = "The total economic damgage on record given \n for the eve
nts with the top 10 highest damage" )
```



Again since there are so many events, we only show the top 10 events in terms of damaged caused. We see from the table and barchart , that floods and hurricanes have caused the most significant amount of economic damage on record, with floods having double the amount of economic damage compared to hurricanes, the cause of the 2nd highest economic damage.