Project2_RepRes

Joerg Heintz

3/16/2017

Summary / Research Question

This investigation provides answers to the following two questions:

1) Across the United States, which types of events are most harmful with respect to population health?
2) Across the United States, which types of events have the greatest economic consequences?

Data Download & Ingestion

```
# set work directory
setwd("/Users/joergheintz/Documents/08_MPHPHI/11_Coursera/Coursera_ReproduciableResearch/RepResProj2")

# download the data set from the source
#mypath<-"https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2"
#download.file(mypath,"myfile.bz2")

mydata<-read.csv("myfile.bz2")[,c("STATE", "EVTYPE", "FATALITIES", "INJURIES", "PROPDMG", "PROPDMGEXP",
```

```
# Split date from time, keep date
BGN_DATE<-data.frame(t(as.data.frame(strsplit(as.character(mydata$BGN_DATE),' ')))[,1], row.names=NULL)
mydata$BGN_DATE<-as.Date(mydata$BGN_DATE, "%m/%d/%Y")


## split data set into Fatalities/Injuiries and EconomicDamage
# Fatalities & Injuries
df_FAT_INJ<-mydata[mydata$FATALITIES != 0 | mydata$INJURIES != 0 ,c("STATE", "EVTYPE", "FATALITIES", "I
write.csv(df_FAT_INJ, "FAT_INJ.csv")


# Economic Damage
df_ECO_DAM<-mydata[mydata$PROPDMG != 0 | mydata$CROPDMG, c("STATE", "EVTYPE", "PROPDMG", "PROPDMGEXP",
write.csv(df_ECO_DAM, "ECO_DAM.csv")
```

Fatilities

The data set is clean from all columns that are not needed, and documented in table.

```
myFAT<-read.csv("FAT_INJ.csv")

# Fatalities Ranking by event Top 10
FAT<-as.data.frame(tapply(myFAT$FATALITIES, myFAT$EVTYPE, sum))
FAT$EVENT<-rownames(FAT)
colnames(FAT)<-c("fatalities", "events")
rownames(FAT)<-NULL
```

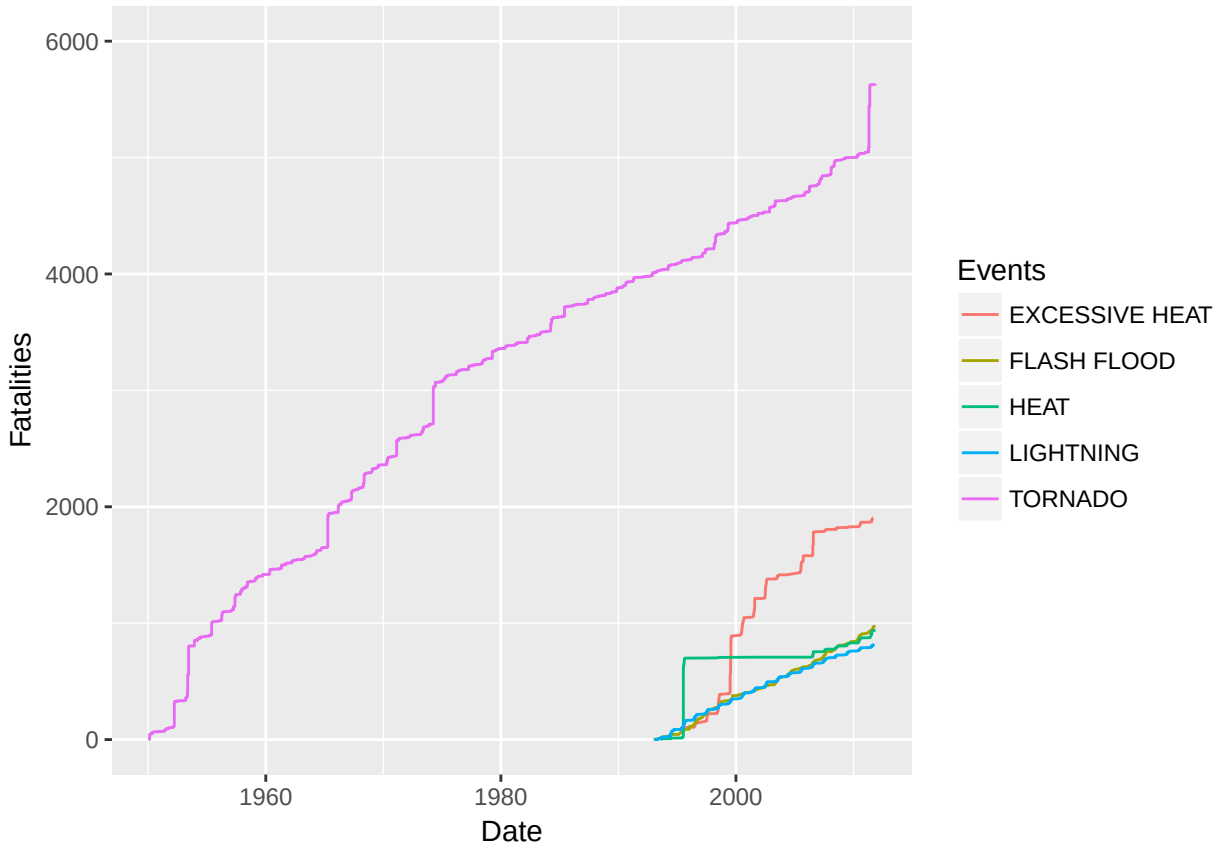
```
FAT<-FAT[order(-FAT$fatalities),]
FAT10<-FAT[1:10,]
FAT10
```

```
##     fatalities         events
## 184       5633        TORNADO
## 32        1903 EXCESSIVE HEAT
## 42         978    FLASH FLOOD
## 69         937           HEAT
## 123        816      LIGHTNING
## 191        504      TSTM WIND
## 47         470          FLOOD
## 147        368    RIP CURRENT
## 93         248      HIGH WIND
## 2          224      AVALANCHE
```

```
FAT5<-FAT[1:5,]
```

```
# Fatalities
# sum up over time
fat<-data.frame()
for (ev in FAT5$events){
        sf<-myFAT[myFAT$EVTYPE == ev, ]
        sf<-sf[order(as.Date(sf$BGN_DATE, format="%Y-%m-%d")),]
        sf$SumFATALITIES<-sf$FATALITIES
        k<-length(sf$FATALITIES)-1
        for (i in 1:k){
                sf$SumFATALITIES[i+1] = sf$SumFATALITIES[i]+sf$SumFATALITIES[i+1]
                }
        k=0
        fat<-rbind(fat,sf)
}
fat$BGN_DATE<-fat$BGN_DATE<-as.Date(fat$BGN_DATE, format="%Y-%m-%d")
fat<-fat[order(as.Date(fat$BGN_DATE, format="%Y-%m-%d")),]
```

```
# Fatalities
g2 = ggplot(fat, aes(x =BGN_DATE , y = SumFATALITIES, group = EVTYPE, color = factor(EVTYPE)))
g2 = g2 + xlab("Date") + ylab("Fatalities") + labs(color = "Events")
g2 = g2 + guides(fill=guide_legend(title="Natural Events"))
g2 = g2 + ylim(0, 6000)
g2 = g2 + geom_line()
g2
```



INJURIES

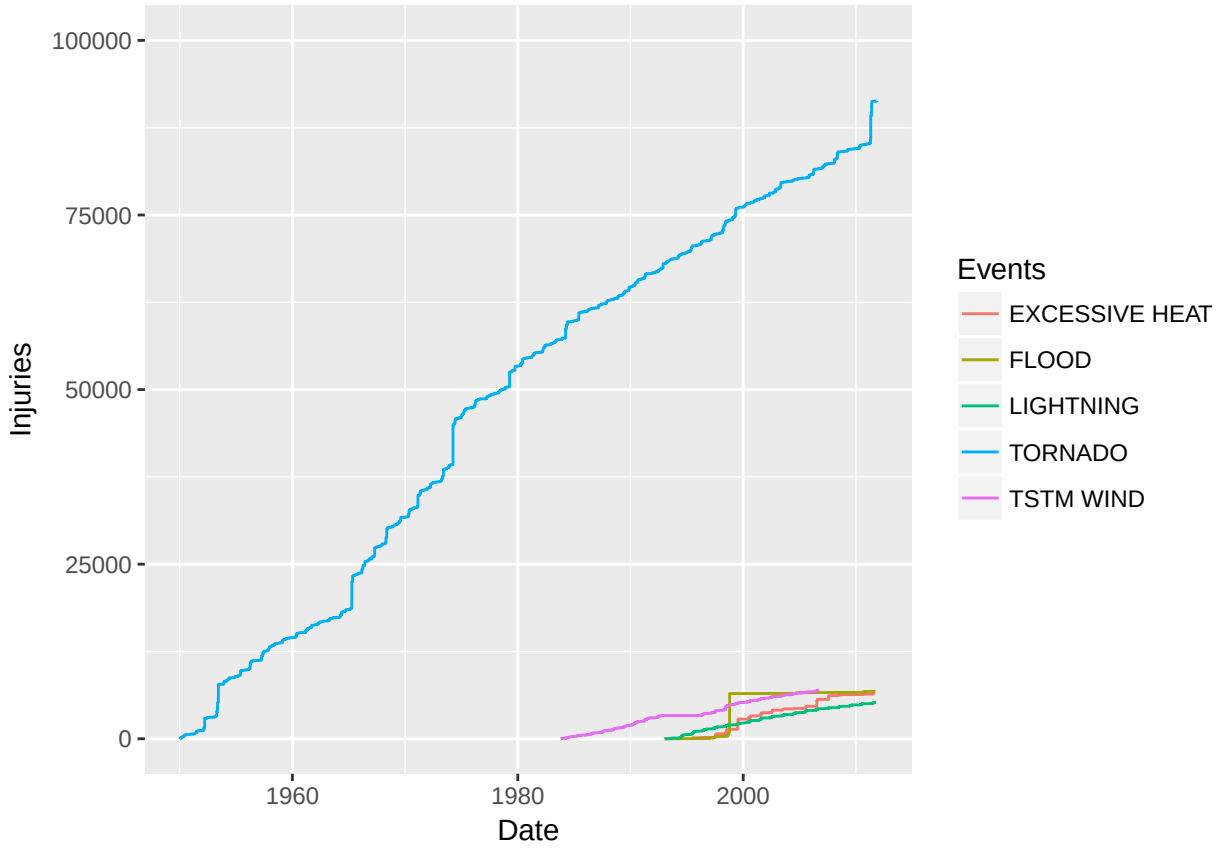
```
myINJ<-read.csv("FAT_INJ.csv")

# Injuries ranking by event Top 10
INJ<-as.data.frame(tapply(myINJ$INJURIES, myINJ$EVTYPE, sum))
INJ$EVENT<-rownames(INJ)
colnames(INJ)<-c("injuries", "events")
rownames(INJ)<-NULL
INJ<-INJ[order(-INJ$injuries),]
INJ10<-INJ[1:10,]
INJ10
```

```
##     injuries           events
## 184    91346          TORNADO
## 191     6957        TSTM WIND
## 47      6789            FLOOD
## 32      6525   EXCESSIVE HEAT
## 123     5230        LIGHTNING
## 69      2100             HEAT
## 117     1975        ICE STORM
## 42      1777      FLASH FLOOD
## 173     1488 THUNDERSTORM WIND
## 67      1361             HAIL
```

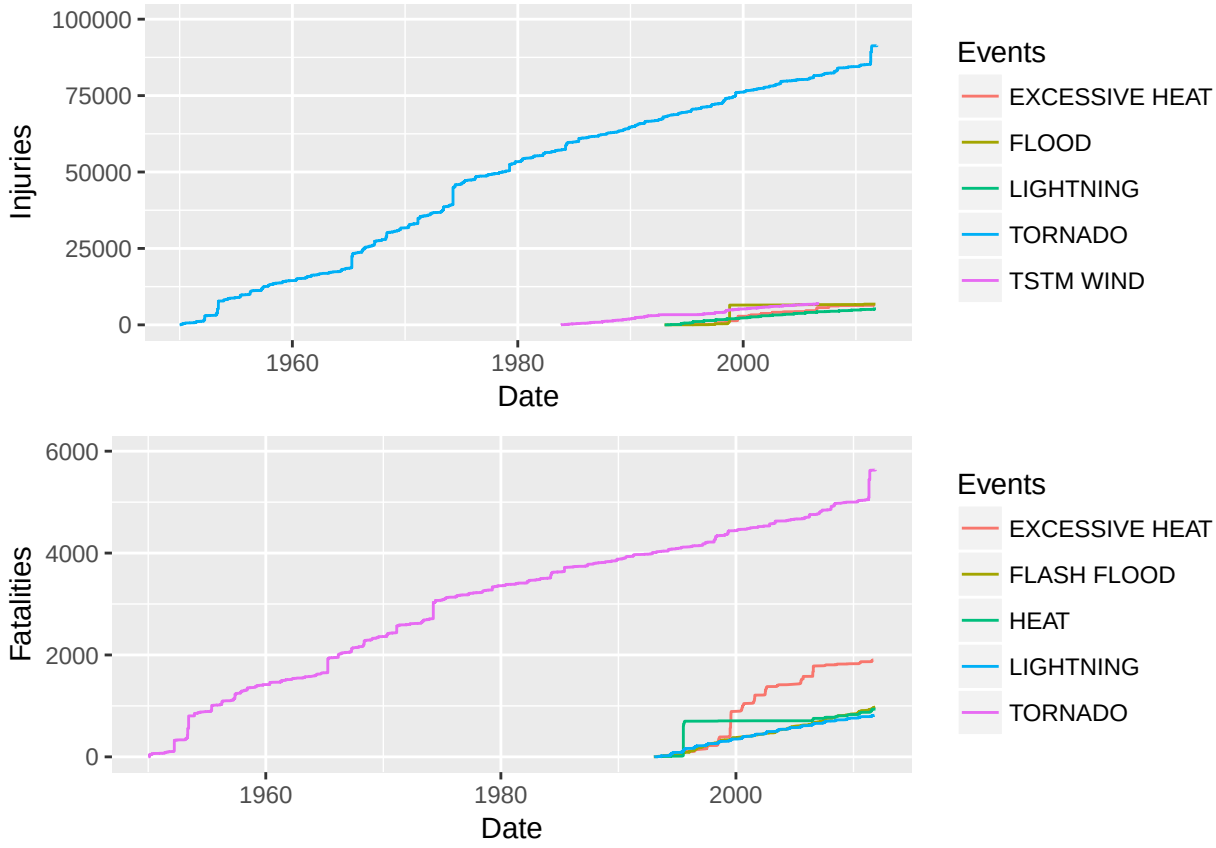
```
INJ5<-INJ[1:5,]
```

```
#Injuries
# sum up over time
inj<-data.frame()
for (ev in INJ5$events){
        sf<-myINJ[myINJ$EVTYPE == ev, ]
        sf<-sf[order(as.Date(sf$BGN_DATE, format="%Y-%m-%d")),]
        sf$SumINJURIES<-sf$INJURIES
        k<-length(sf$INJURIES)-1
        for (i in 1:k){
                sf$SumINJURIES[i+1] = sf$SumINJURIES[i]+sf$SumINJURIES[i+1]
                }
        k=0
        inj<-rbind(inj,sf)
}
inj$BGN_DATE<-inj$BGN_DATE<-as.Date(inj$BGN_DATE, format="%Y-%m-%d")
inj<-inj[order(as.Date(inj$BGN_DATE, format="%Y-%m-%d")),]
```

```
# Injuries
g1 = ggplot(inj, aes(x =BGN_DATE ,y = SumINJURIES,  color = factor(EVTYPE)))
g1 = g1 + guides(fill=guide_legend(title="Natural Events")) + labs(color = "Events")
g1 = g1 + xlab("Date") + ylab("Injuries")
g1 = g1 + ylim(0, 100000)
g1 = g1 + geom_line()
g1
```



```
# Combining fatalities and injuries in one plot
pFat<-ggplotGrob(g2)
pInf<-ggplotGrob(g1)
grid.arrange(pFat, pInf, layout_matrix = rbind(2,1))
```



Economic Damages

```
myDAM<-read.csv("ECO_DAM.csv")

# Property Damage
myDAM$PROPDMGEXP<-as.character(myDAM$PROPDMGEXP)
myDAM[myDAM$PROPDMGEXP != as.character("K") & myDAM$PROPDMGEXP != as.character("M") & myDAM$PROPDMGEXP
myDAM[myDAM$PROPDMGEXP == "K", "PROPDMGEXP"] <- 1000
myDAM[myDAM$PROPDMGEXP == "M", "PROPDMGEXP"] <- 1000000
myDAM[myDAM$PROPDMGEXP == "B", "PROPDMGEXP"] <- 1000000000
myDAM$PROPDMGEXP<-as.numeric(myDAM$PROPDMGEXP)
myDAM$PROPDMG <- as.numeric(myDAM$PROPDMG) * as.numeric(myDAM$PROPDMGEXP)
```

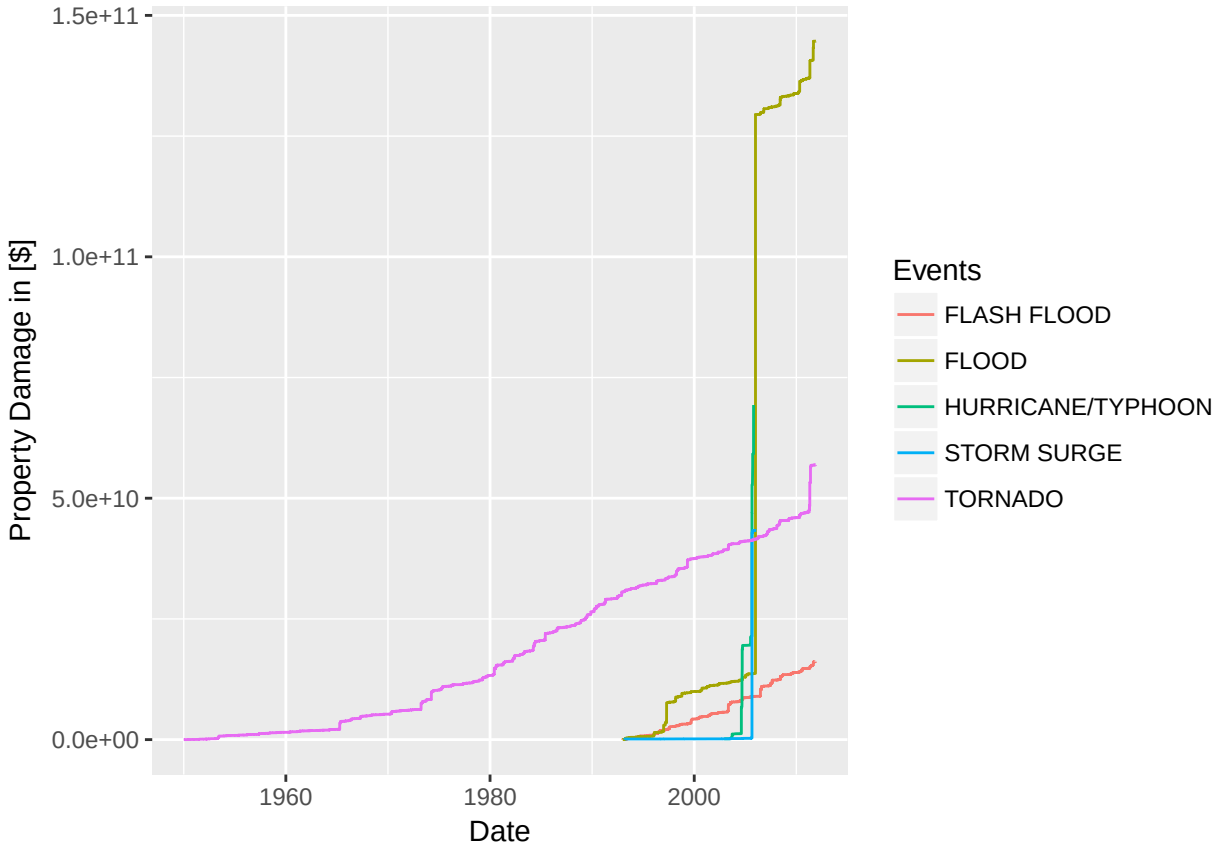
```
# property damage, sum up damages deriving top events
DamProp<-as.data.frame(tapply(myDAM$PROPDMG, myDAM$EVTYPE, sum))
DamProp$EVENT<-rownames(DamProp)
colnames(DamProp)<-c("damage", "events")
rownames(DamProp)<-NULL
DamProp<-DamProp[order(-DamProp$damage),]
DamProp[1:10,]
```

```
##           damage            events
## 72  144657709800             FLOOD
## 197  69305840000 HURRICANE/TYPHOON
## 354  56925660480           TORNADO
## 299  43323536000       STORM SURGE
## 59   16140811510       FLASH FLOOD
## 116  15727366720              HAIL
## 189  11868319010         HURRICANE
## 363   7703890550    TROPICAL STORM
## 424   6688497250      WINTER STORM
## 174   5270046260         HIGH WIND
```

```
TopPropDam<-DamProp[1:5,]
```

```
#Property damage, sum up over time
propDam<-data.frame()
for (ev in TopPropDam$events){
        sf<-myDAM[myDAM$EVTYPE == ev, ]
        sf<-sf[order(as.Date(sf$BGN_DATE, format="%Y-%m-%d")),]
        sf$SummyDAM<-sf$PROPDMG
        k<-length(sf$PROPDMG)-1
        for (i in 1:k){
                sf$SummyDAM[i+1] = sf$SummyDAM[i]+sf$SummyDAM[i+1]
                }
        k=0
        propDam<-rbind(propDam,sf)
}

propDam$BGN_DATE<-as.Date(propDam$BGN_DATE, format="%Y-%m-%d")
propDam<-propDam[order(as.Date(propDam$BGN_DATE, format="%Y-%m-%d")),]
```

```
#Property Damage
g3 = ggplot(propDam, aes(x =BGN_DATE , y = SummyDAM, group = EVTYPE, color = factor(EVTYPE)))
g3 = g3 + xlab("Date") + ylab("Property Damage in [$]")+ labs(color = "Events")
g3 = g3 + geom_line()
g3
```



Crop Damage

```
myDAM<-read.csv("ECO_DAM.csv")

#Crop Damage
myDAM$CROPDMGEXP<-as.character(myDAM$CROPDMGEXP)
myDAM[myDAM$CROPDMGEXP != as.character("K") & myDAM$CROPDMGEXP != as.character("M") & myDAM$CROPDMGEXP
myDAM[myDAM$CROPDMGEXP == "K", "CROPDMGEXP"] <- 1000
myDAM[myDAM$CROPDMGEXP == "M", "CROPDMGEXP"] <- 1000000
myDAM[myDAM$CROPDMGEXP == "B", "CROPDMGEXP"] <- 1000000000
myDAM$CROPDMG <- as.numeric(myDAM$CROPDMG) * as.numeric(myDAM$CROPDMGEXP)
```

```
#sum up damages and derivng top events
DamCrop<-as.data.frame(tapply(myDAM$CROPDMG, myDAM$EVTYPE, sum))
DamCrop$EVENT<-rownames(DamCrop)
colnames(DamCrop)<-c("damage", "events")
rownames(DamCrop)<-NULL
DamCrop<-DamCrop[order(-DamCrop$damage),]
DamCrop[1:10,]
```

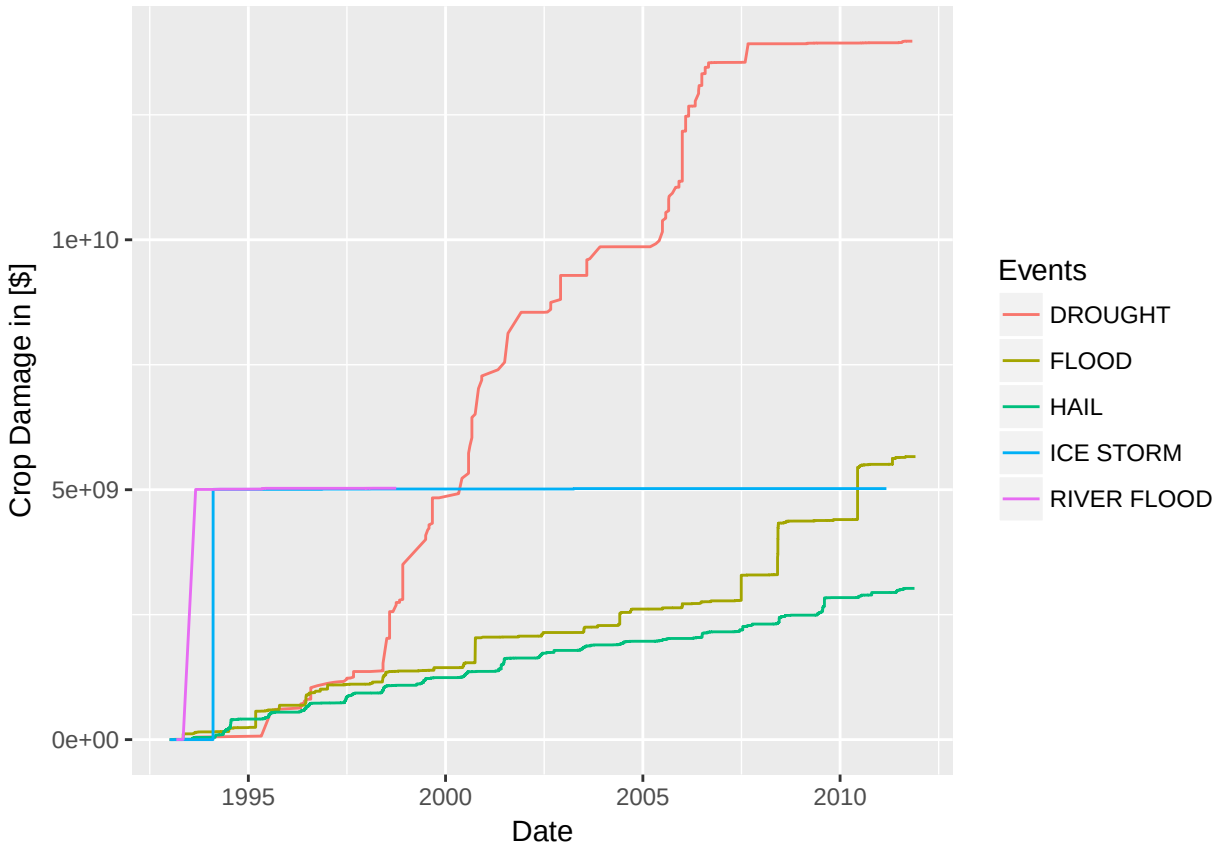
```
##          damage            events
## 39  13972566000           DROUGHT
## 72   5661968450             FLOOD
## 262  5029459000       RIVER FLOOD
## 206  5022113500         ICE STORM
```

```
## 116  3025537450              HAIL
## 189  2741910000         HURRICANE
## 197  2607872800 HURRICANE/TYPHOON
## 59   1421317100       FLASH FLOOD
## 54   1292973000      EXTREME COLD
## 98   1094086000      FROST/FREEZE
```

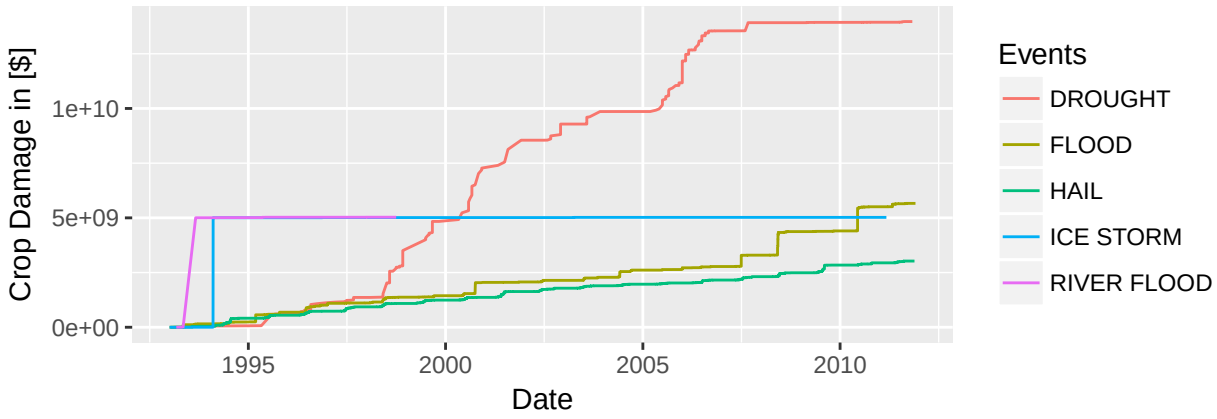
```
TopCropDam<-DamCrop[1:5,]
```

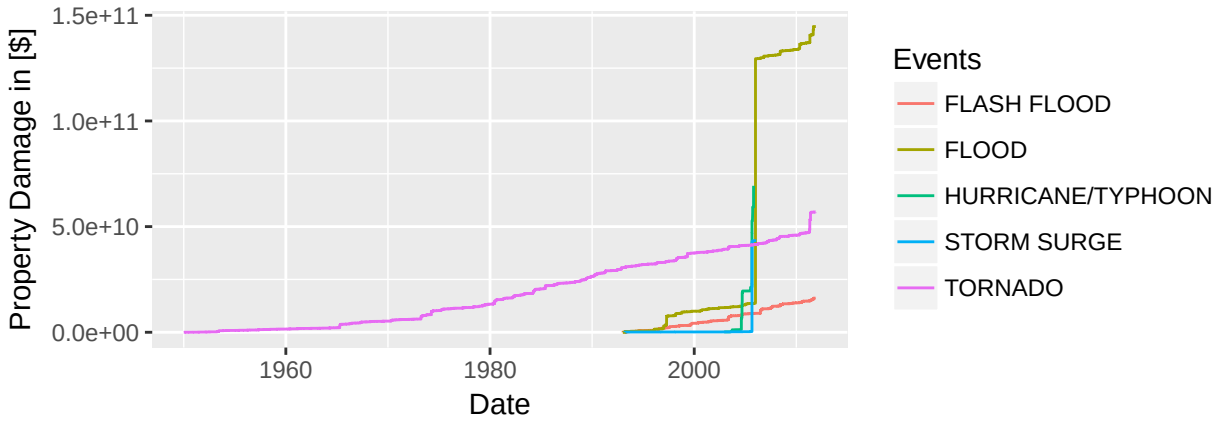
```
#Crop damage, sum up over time
cropDam<-data.frame()
for (ev in TopCropDam$events){
        sf<-myDAM[myDAM$EVTYPE == ev, ]
        sf<-sf[order(as.Date(sf$BGN_DATE, format="%Y-%m-%d")),]
        sf$SummyCropDAM<-sf$CROPDMG
        k<-length(sf$CROPDMG)-1
        for (i in 1:k){
                sf$SummyCropDAM[i+1] = sf$SummyCropDAM[i]+sf$SummyCropDAM[i+1]
                }
        k=0
        cropDam<-rbind(cropDam,sf)
}

cropDam$BGN_DATE<-as.Date(cropDam$BGN_DATE, format="%Y-%m-%d")
cropDam<-cropDam[order(as.Date(cropDam$BGN_DATE, format="%Y-%m-%d")),]
```

```
#Crop Damage
g4 = ggplot(cropDam, aes(x =BGN_DATE , y = SummyCropDAM, group = EVTYPE, color = factor(EVTYPE)))
g4 = g4 + xlab("Date") + ylab("Crop Damage in [$]")+ labs(color = "Events")
g4 = g4 + geom_line()
g4
```




```
# Combining property and crop damages
pProp<-ggplotGrob(g3)
pCrop<-ggplotGrob(g4)
grid.arrange(pProp, pCrop, layout_matrix = rbind(2,1))
```


Crop Damage in [$]
1e+10
5e+09
0e+00
1995
2000
2005
2010
Date
Events
DROUGHT
FLOOD
HAIL
ICE STORM
RIVER FLOOD


Property Damage in [$]
1.5e+11
1.0e+11
5.0e+10
0.0e+00
1960
1980
2000
Date
Events
FLASH FLOOD
FLOOD
HURRICANE/TYPHOON
STORM SURGE
TORNADO