Analyzing the NOAA Storm Database

Synopsis

The basic goal of this assignment is to explore the NOAA Storm Database which contains events from 1950 to Nov 2011 and answer some basic questions about severe weather events. Will try to analyze & present my findings in the following areas:

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?

Introduction

Storms and other severe weather events can cause both public health and economic problems for communities and municipalities. Many severe events can result in fatalities, injuries, and property damage, and preventing such outcomes to the extent possible is a key concern.

This project involves exploring 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, and property damage.

Data

The stormData for this assignment come in the form of a comma-separated-value file compressed via the bzip2 algorithm to reduce its size. You can download the file from the course web site:

Storm Data[47Mb]

There is also some documentation of the Database available. Here you will find how some of the variables are constructed/defined.

1. National Weather Service Storm Data Documentation
2. National Climatic stormData Center Storm Events FAQ

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 records. More recent years should be considered more complete.

Data Processing

Set the working directory

```
setwd(".")
```

Download the file

```
download.file("https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2",
destfile = "./stormData.csv.bz2",method = "curl")
```

Read stormData from the downloaded file

```
stormData <- read.csv("./stormData.csv.bz2")

dim(stormData)

## [1] 902297     37
```

Calculations & Data transformations

Load required packages

```
library(ggplot2)

library(plyr)
```

Finding the total Harm with sum of FATALITIES and INJURIES by EVTYPE

```
injuryDataFrame <- ddply(stormData, .(EVTYPE), summarize, TotalHarm = sum(FATALITIES + IN
JURIES))

injuryDataFrame <- injuryDataFrame[order(injuryDataFrame$TotalHarm, decreasing = T), ]
```

Top 10 Harm

```
TopHarm <- injuryDataFrame[1:10, ]
```

Property Damage: Find the sum of PROPDMG by EVTYPE and PROPDMGEXP.

```
prop <- ddply(stormData, .(EVTYPE, PROPDMGEXP), summarize, PROPDMG = sum(PROPDMG))
```

Property Damage: Finding the value of property Damage

```
prop <- mutate(prop, PropertyDamage = ifelse(toupper(PROPDMGEXP) =='K', PROPDMG*1000, ife
lse(toupper(PROPDMGEXP) =='M', PROPDMG*1000000, ifelse(toupper(PROPDMGEXP) == 'B', PROPDM
G*1000000000, ifelse(toupper(PROPDMGEXP) == 'H', PROPDMG*100, PROPDMG)))))
```

Property Damage: Finding the property damage based on event type

```
prop <- subset(prop, select = c("EVTYPE", "PropertyDamage"))

prop.total <- ddply(prop, .(EVTYPE), summarize, TotalPropDamage = sum(PropertyDamage))
```

Crop Damage: Sum of the the CROPDMG by EVTYPE and CROPDMGEXP.

```
crop <- ddply(stormData, .(EVTYPE, CROPDMGEXP), summarize, CROPDMG = sum(CROPDMG))
```

Crop Damage: Real crop damage based on CROPDMGEXP.

```
crop <- mutate(crop, CropDamage = ifelse(toupper(CROPDMGEXP) =='K', CROPDMG*1000, ifelse(
toupper(CROPDMGEXP) =='M', CROPDMG*1000000, ifelse(toupper(CROPDMGEXP) == 'B', CROPDMG*10
00000000, ifelse(toupper(CROPDMGEXP) == 'H', CROPDMG*100, CROPDMG)))))
```

Crop Damage: Sum of the crop damage by event type

```
crop <- subset(crop, select = c("EVTYPE", "CropDamage"))
```

```
crop.total <- ddply(crop, .(EVTYPE), summarize, TotalCropDamage = sum(CropDamage))
```

Total Damage : Merging the Property & Crop Damage

```
damageDataFrame <- merge(prop.total, crop.total, by="EVTYPE")

damageDataFrame <- mutate(damageDataFrame, TotalDamage = TotalPropDamage + TotalCropDamag
e)

damageDataFrame <- damageDataFrame[order(damageDataFrame$TotalDamage, decreasing = T), ]
```

Top 10 Damage

```
TopDamage <- damageDataFrame[1:10, ]
```

Results

1 . Population Health Casualties

This is the result of top 10 harmful types base on the sum of casualties.

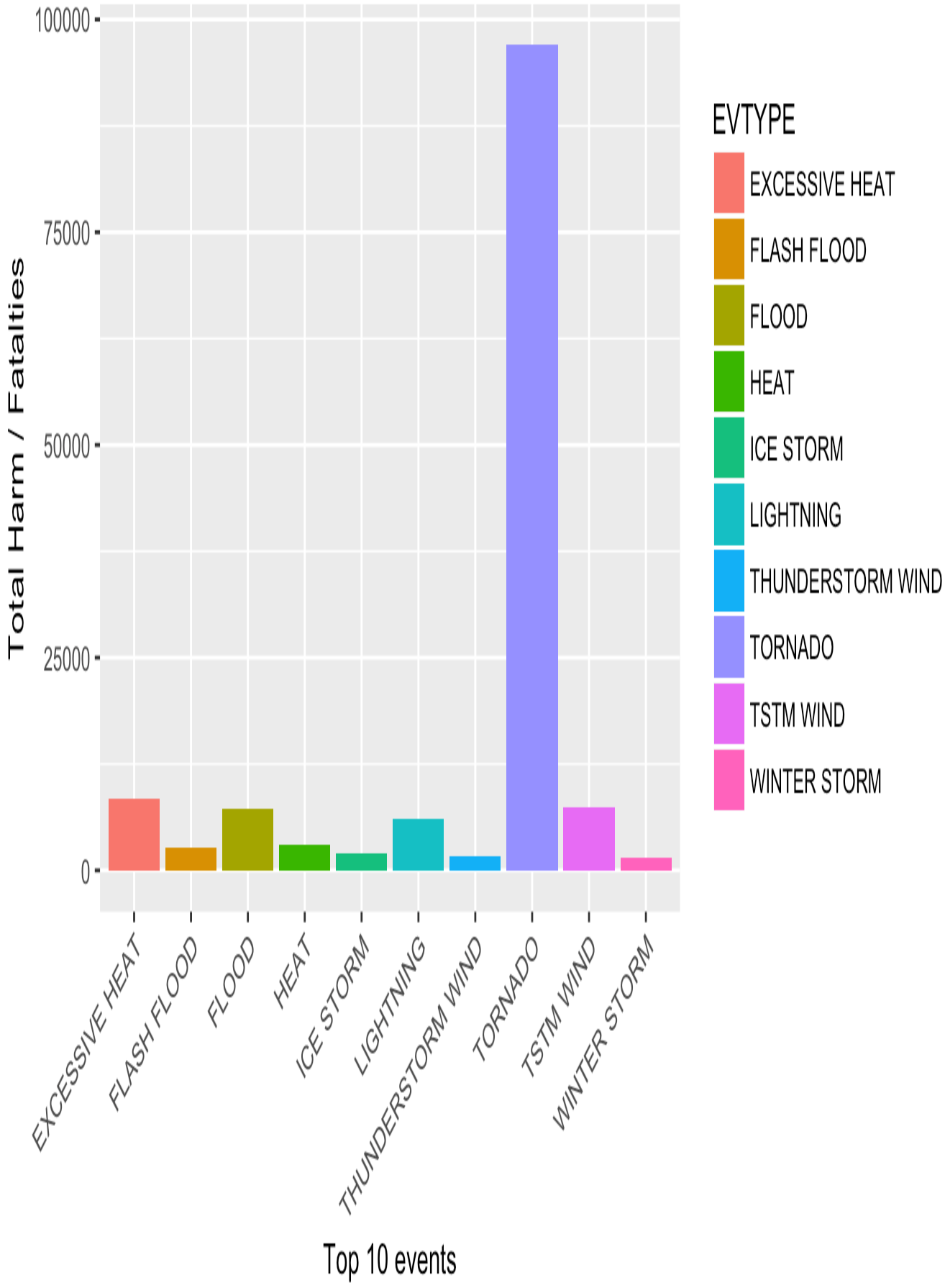
```
TopHarm
```

```
##                EVTYPE TotalHarm
## 834           TORNADO     96979
## 130    EXCESSIVE HEAT      8428
## 856         TSTM WIND      7461
## 170             FLOOD      7259
## 464         LIGHTNING      6046
## 275              HEAT      3037
## 153       FLASH FLOOD      2755
## 427         ICE STORM      2064
## 760 THUNDERSTORM WIND      1621
## 972      WINTER STORM      1527
```

This is the plot base on previous stormData

```
totalHarmPlot <- ggplot(TopHarm, aes( EVTYPE,TotalHarm, fill=EVTYPE)) + geom_bar(stat="id
entity") + xlab("Top 10 events")+ ylab("Total Harm / Fatalties")+ ggtitle("Fatalities due
to severe weather events in the U.S from 1950-2011") + theme(axis.text.x=element_text(ang
le=45,hjust=1))

totalHarmPlot
```


Fatalities due to severe weather events in the U.S from 1950-2011
Total Harm / Fatalties
100000
75000
50000
25000
0
EXCESSIVE HEAT
FLASH FLOOD
FLOOD
HEAT
ICE STORM
LIGHTNING
THUNDERSTORM WIND
TORNADO
TSTM WIND
WINTER STORM
Top 10 events
EVTYPE
EXCESSIVE HEAT
FLASH FLOOD
FLOOD
HEAT
ICE STORM
LIGHTNING
THUNDERSTORM WIND
TORNADO
TSTM WIND
WINTER STORM

Most fatalties are caused by Tornado

2. Economic Casulaties

Here are the top 10 damages caused

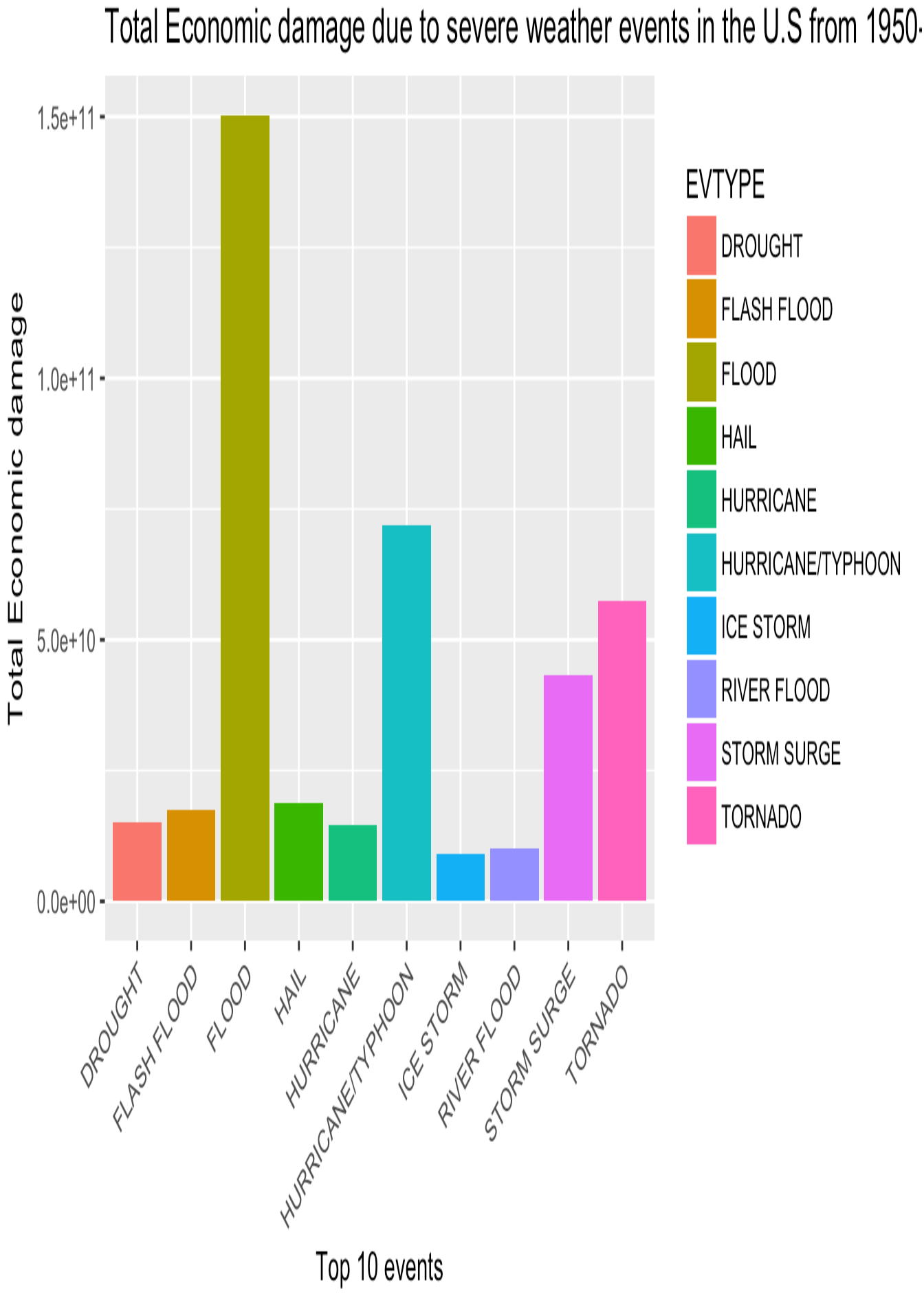
```
TopDamage
```

```
##                EVTYPE TotalPropDamage TotalCropDamage  TotalDamage
## 170             FLOOD    144657709807      5661968450 150319678257
## 411 HURRICANE/TYPHOON     69305840000      2607872800  71913712800
## 834           TORNADO     56937160779       414953270  57352114049
## 670       STORM SURGE     43323536000            5000  43323541000
## 244              HAIL     15732267543      3025954473  18758222016
## 153       FLASH FLOOD     16140812067      1421317100  17562129167
## 95            DROUGHT      1046106000     13972566000  15018672000
## 402         HURRICANE     11868319010      2741910000  14610229010
## 590       RIVER FLOOD      5118945500      5029459000  10148404500
## 427         ICE STORM      3944927860      5022113500   8967041360
```

This is the plot base on the Total damage : sum of totalCropDamage & totalPropDamage

```
totaldamagePlot <- ggplot(TopDamage, aes( EVTYPE,TotalDamage, fill=EVTYPE)) + geom_bar(st
at="identity") + xlab("Top 10 events")+ ylab("Total Economic damage")+ ggtitle("Total Eco
nomic damage due to severe weather events in the U.S from 1950-2011") + theme(axis.text.x
=element_text(angle=45,hjust=1))

totaldamagePlot
```

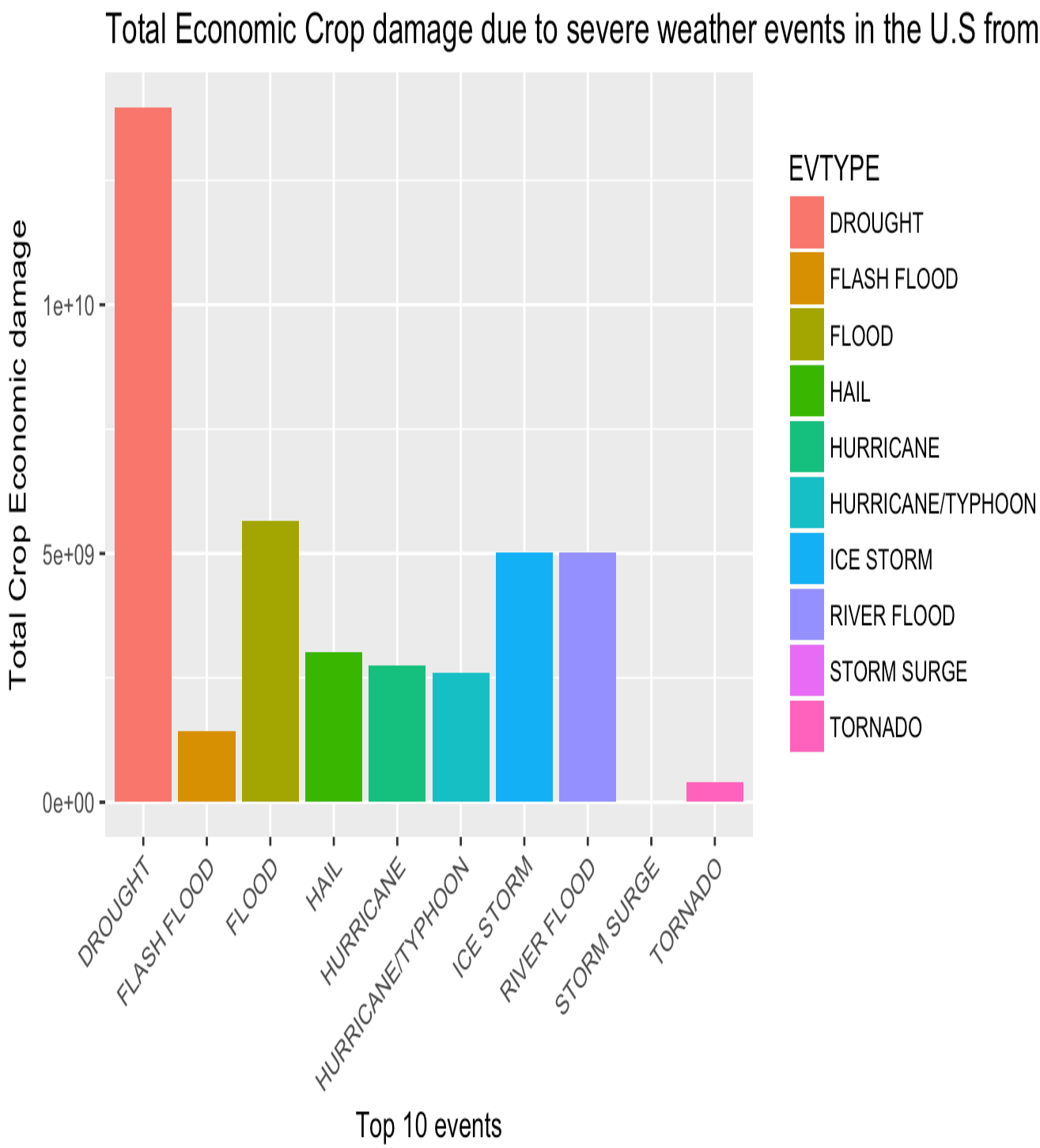

Total Economic damage due to severe weather events in the U.S from 1950-
Total Economic damage
1.5e+11
1.0e+11
5.0e+10
0.0e+00
DROUGHT
FLASH FLOOD
FLOOD
HAIL
HURRICANE
HURRICANE/TYPHOON
ICE STORM
RIVER FLOOD
STORM SURGE
TORNADO
Top 10 events
EVTYPE
DROUGHT
FLASH FLOOD
FLOOD
HAIL
HURRICANE
HURRICANE/TYPHOON
ICE STORM
RIVER FLOOD
STORM SURGE
TORNADO

Most damages are caused by: Flood

This is the plot based on the Total CropDamage

```
totalcropDamagePlot <- ggplot(TopDamage, aes( EVTYPE,TotalCropDamage, fill=EVTYPE)) + geo
m_bar(stat="identity") + xlab("Top 10 events")+ ylab("Total Crop Economic damage")+ ggtit
le("Total Economic Crop damage due to severe weather events in the U.S from 1950-2011") +
theme(axis.text.x=element_text(angle=45,hjust=1))

totalcropDamagePlot
```



Most Crop Damages are caused by : Drought