Reproducible Research Week 4:Course Project 2

Ravinendra Pratap

4 August 2019

R Markdown

Reproducible Research: Peer-graded Assignment: Course Project 2

Introduction:

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.

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.

Synopsis:

The basic goal of this assignment is to explore the NOAA Storm Database and answer two basic questions about severe weather events. First one, the types of events are most harmful to population health and secondly the types of events have the greatest economic consequences.

Analysis:

The Analysis on the storm event database revealed that tornadoes are the most harmful weather event to the population's health. The second most harmful event type is excessive heat. The greatest economic impact of weather events were also analysed. Flash floods and thunderstorm winds caused billions of dollars in property damages between 1950 and 2011. The largest damage to crops were caused by droughts, followed by floods and hailing.

1. Load Data

1.Loading packages (rmarkdown, knitr and dplyr)

```
library(rmarkdown)
```

```
## Warning: package 'rmarkdown' was built under R version 3.5.3
```

```
library(knitr)
```

```
## Warning: package 'knitr' was built under R version 3.5.3
```

```
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 3.5.3
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

2.Getting Working Directory

```
getwd()
```

```
## [1] "D:/LND/COURSERA_DATA_SCIENCE/COURSERA_05_Reproducible Research/WEEK4_05RR_Markdown_knitr"
```

```
setwd ("D:/LND/COURSERA_DATA_SCIENCE/COURSERA_05_Reproducible Research/WEEK4_05RR_Markdown_knitr")
getwd()
```

```
## [1] "D:/LND/COURSERA_DATA_SCIENCE/COURSERA_05_Reproducible Research/WEEK4_05RR_Markdown_knitr"
```

3.Load the data (i.e. read.csv())
4.Downloading zip file if it doesn't already exist in the workspace
5.Clear the workspace load raw activity data
6.Read the dataset and store in "StormData". View columns afterwards

```
#url <- "https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2"
#download.file(url, "StormData.csv.bz2", mode='wb')
#rm(list=ls())
StormData <- read.csv(bzfile("StormData.csv.bz2"))
names(StormData)
```

```
##  [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"
```

2. Data Processing

1. Choose the columns and store in "StormData2". In this project we only use the following ("BGN_DATE","EVTYPE","FATALITIES","INJURIES","PROPDMG","PROPDMGEXP","CROPDMG","CROPDMGEXP") Read the dataset.

```
StormData2 <-  StormData[ ,c( "BGN_DATE","EVTYPE","FATALITIES","INJURIES","PROPDMG",
"PROPDMGEXP","CROPDMG","CROPDMGEXP")]
```

2.View Top and Bottom rows of StormData2

```
head (StormData2,10)
```

```
##             BGN_DATE  EVTYPE FATALITIES INJURIES PROPDMG PROPDMGEXP
## 1   4/18/1950 0:00:00 TORNADO          0       15    25.0          K
## 2   4/18/1950 0:00:00 TORNADO          0        0     2.5          K
## 3   2/20/1951 0:00:00 TORNADO          0        2    25.0          K
## 4    6/8/1951 0:00:00 TORNADO          0        2     2.5          K
## 5  11/15/1951 0:00:00 TORNADO          0        2     2.5          K
## 6  11/15/1951 0:00:00 TORNADO          0        6     2.5          K
## 7  11/16/1951 0:00:00 TORNADO          0        1     2.5          K
## 8   1/22/1952 0:00:00 TORNADO          0        0     2.5          K
## 9   2/13/1952 0:00:00 TORNADO          1       14    25.0          K
## 10  2/13/1952 0:00:00 TORNADO          0        0    25.0          K
##    CROPDMG CROPDMGEXP
## 1        0
## 2        0
## 3        0
## 4        0
## 5        0
## 6        0
## 7        0
## 8        0
## 9        0
## 10       0
```

```
tail (StormData2,10)
```

```
##                 BGN_DATE         EVTYPE FATALITIES INJURIES PROPDMG
## 902288  11/5/2011 0:00:00 WINTER WEATHER          0        0       0
## 902289 11/28/2011 0:00:00   FROST/FREEZE          0        0       0
## 902290 11/12/2011 0:00:00      HIGH WIND          0        0       0
## 902291 11/28/2011 0:00:00 WINTER WEATHER          0        0       0
## 902292 11/28/2011 0:00:00 WINTER WEATHER          0        0       0
## 902293 11/30/2011 0:00:00      HIGH WIND          0        0       0
## 902294 11/10/2011 0:00:00      HIGH WIND          0        0       0
## 902295  11/8/2011 0:00:00      HIGH WIND          0        0       0
## 902296  11/9/2011 0:00:00       BLIZZARD          0        0       0
## 902297 11/28/2011 0:00:00     HEAVY SNOW          0        0       0
##        PROPDMGEXP CROPDMG CROPDMGEXP
## 902288          K       0          K
## 902289          K       0          K
## 902290          K       0          K
## 902291          K       0          K
## 902292          K       0          K
## 902293          K       0          K
## 902294          K       0          K
## 902295          K       0          K
## 902296          K       0          K
## 902297          K       0          K
```

3.Check the unique values of PROPDMGEX (Property Damage Expences)

```
unique(StormData2$PROPDMGEXP)
```

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

4. Convert values of PROPDMGEXP to the followin

- “H” or “h” = 2
- “M” or “m” = 6

- “K” = 3
- “B” = 9
- “+”, “-” , “?” , “” = 0 In converting convert first the column into character

```
StormData2$PROPDMGEXP <- as.character(StormData2$PROPDMGEXP)
StormData2$PROPDMGEXP[StormData2$PROPDMGEXP == "B"] <- 9
StormData2$PROPDMGEXP[StormData2$PROPDMGEXP == "K"] <- 3
StormData2$PROPDMGEXP[StormData2$PROPDMGEXP == "M"] <- 6
StormData2$PROPDMGEXP[StormData2$PROPDMGEXP == "m"] <- 6
StormData2$PROPDMGEXP[StormData2$PROPDMGEXP == "h"] <- 2
StormData2$PROPDMGEXP[StormData2$PROPDMGEXP == "H"] <- 2
StormData2$PROPDMGEXP[StormData2$PROPDMGEXP == "+"] <- 0
StormData2$PROPDMGEXP[StormData2$PROPDMGEXP == "?"] <- 0
StormData2$PROPDMGEXP[StormData2$PROPDMGEXP == "-"] <- 0
StormData2$PROPDMGEXP[StormData2$PROPDMGEXP == ""] <- 0
```

5.Do the same to column CROPDMGEXP (Crop Damage Expences)

```
unique(StormData2$CROPDMGEXP)
```

```
## [1]   M K m B ? 0 k 2
## Levels:  ? 0 2 B k K m M
```

```
StormData2$CROPDMGEXP <- as.character(StormData2$CROPDMGEXP)
StormData2$CROPDMGEXP[StormData2$CROPDMGEXP == "B"] <- 9
StormData2$CROPDMGEXP[StormData2$CROPDMGEXP == "K"] <- 3
StormData2$CROPDMGEXP[StormData2$CROPDMGEXP == "k"] <- 3
StormData2$CROPDMGEXP[StormData2$CROPDMGEXP == "M"] <- 6
StormData2$CROPDMGEXP[StormData2$CROPDMGEXP == "m"] <- 6
StormData2$CROPDMGEXP[StormData2$CROPDMGEXP == "?"] <- 0
StormData2$CROPDMGEXP[StormData2$CROPDMGEXP == ""] <- 0

unique(StormData2$CROPDMGEXP)
```

```
## [1] "0" "6" "3" "9" "2"
```

6. Now we need to add a column and compute total Crop Damage and total Property Damage. Prior to computing we need to convert the column to Numeric.

```
StormData2$PROPDMGEXP <- as.numeric(StormData2$PROPDMGEXP)
StormData2$PROPTOTAL <- StormData2$PROPDMG * 10 ^ StormData2$PROPDMGEXP
StormData2$CROPDMGEXP <- as.numeric(StormData2$CROPDMGEXP)
StormData2$CROPTOTAL <- StormData2$CROPDMG * 10 ^ StormData2$CROPDMGEXP
StormData2$ECODMG <- StormData2$PROPTOTAL + StormData2$CROPTOTAL
```

3. Data Analysis

Now that we have clean data we will now aggregate the columns INJURIES, FATALITIES, and ECODMG by EVTYPE

By aggregating we will determine the Event Type that has the most Injuries, Fatalities, and ECODMG

In determining the effect of the Event Type in population health we need to break it down to two which are Injuries and Fatalities
we need to Aggregate FATALITIES and INJURIES by EVTYPE by computing its sum then we will need to get the top 10 of it.
(Choosing the top 10 will reduce the variable in order to see the top most Event that has the greatest efect)

1. Now we need to aggregate FATALITIES

```
StormData2.aggfatalities <- aggregate(FATALITIES ~ EVTYPE, data = StormData2, FUN = "sum")
StormData2.top10fatalities<- StormData2.aggfatalities[order(-StormData2.aggfatalities$FATALITIES),][1:
10,]
```

2. View Top 10 FATALITIES

```
head(StormData2.top10fatalities,10)
```

```
##             EVTYPE FATALITIES
## 834        TORNADO       5633
## 130 EXCESSIVE HEAT       1903
## 153    FLASH FLOOD        978
## 275           HEAT        937
## 464      LIGHTNING        816
## 856      TSTM WIND        504
## 170          FLOOD        470
## 585    RIP CURRENT        368
## 359      HIGH WIND        248
## 19       AVALANCHE        224
```

OR Instead of above Syntax can use below
StormData2.fatalities <- StormData2 %>% select(EVTYPE, FATALITIES) %>% group_by(EVTYPE) %>%
summarise(total.fatalities = sum(FATALITIES)) %>% arrange(-total.fatalities)
head(StormData2.fatalities,10)

3. Now we need to aggregate INJURIES

```
StormData2.agginjuries <- aggregate(INJURIES ~ EVTYPE, data = StormData2, FUN = "sum")
StormData2.top10injuries <- StormData2.agginjuries[order(-StormData2.agginjuries$INJURIES),][1:10,]
```

4. View Top 10 INJURIES

```
head(StormData2.top10injuries,10)
```

```
##               EVTYPE INJURIES
## 834           TORNADO    91346
## 856         TSTM WIND     6957
## 170             FLOOD     6789
## 130    EXCESSIVE HEAT     6525
## 464         LIGHTNING     5230
## 275              HEAT     2100
## 427         ICE STORM     1975
## 153       FLASH FLOOD     1777
## 760 THUNDERSTORM WIND     1488
## 244              HAIL     1361
```

OR Instead of above Syntax can use below
StormData2.injuries <- StormData2 %>% select(EVTYPE, INJURIES) %>% group_by(EVTYPE) %>%
summarise(total.injuries = sum(INJURIES)) %>% arrange(-total.injuries)
head(StormData2.injuries, 10)
tail(StormData2.injuries, 10)

5. Calculate TOTAL Health

```
StormData2.aggeconomic <- aggregate(ECODMG ~ EVTYPE, data = StormData2, FUN = "sum")
StormData2.top10eco <- StormData2.aggeconomic[order(-StormData2.aggeconomic$ECODMG),][1:10,]
head(StormData2.top10eco ,10)
```

```
##                EVTYPE       ECODMG
## 170             FLOOD 150319678257
## 411 HURRICANE/TYPHOON  71913712800
## 834           TORNADO  57362333947
## 670       STORM SURGE  43323541000
## 244              HAIL  18761221986
## 153       FLASH FLOOD  18243991079
## 95            DROUGHT  15018672000
## 402         HURRICANE  14610229010
## 590       RIVER FLOOD  10148404500
## 427         ICE STORM   8967041360
```

4. Results and Discussion

For determining the result we will now graph the aggregated top 10 of each categories (Fatalities, Injuries, Economic Damage) by Event Type

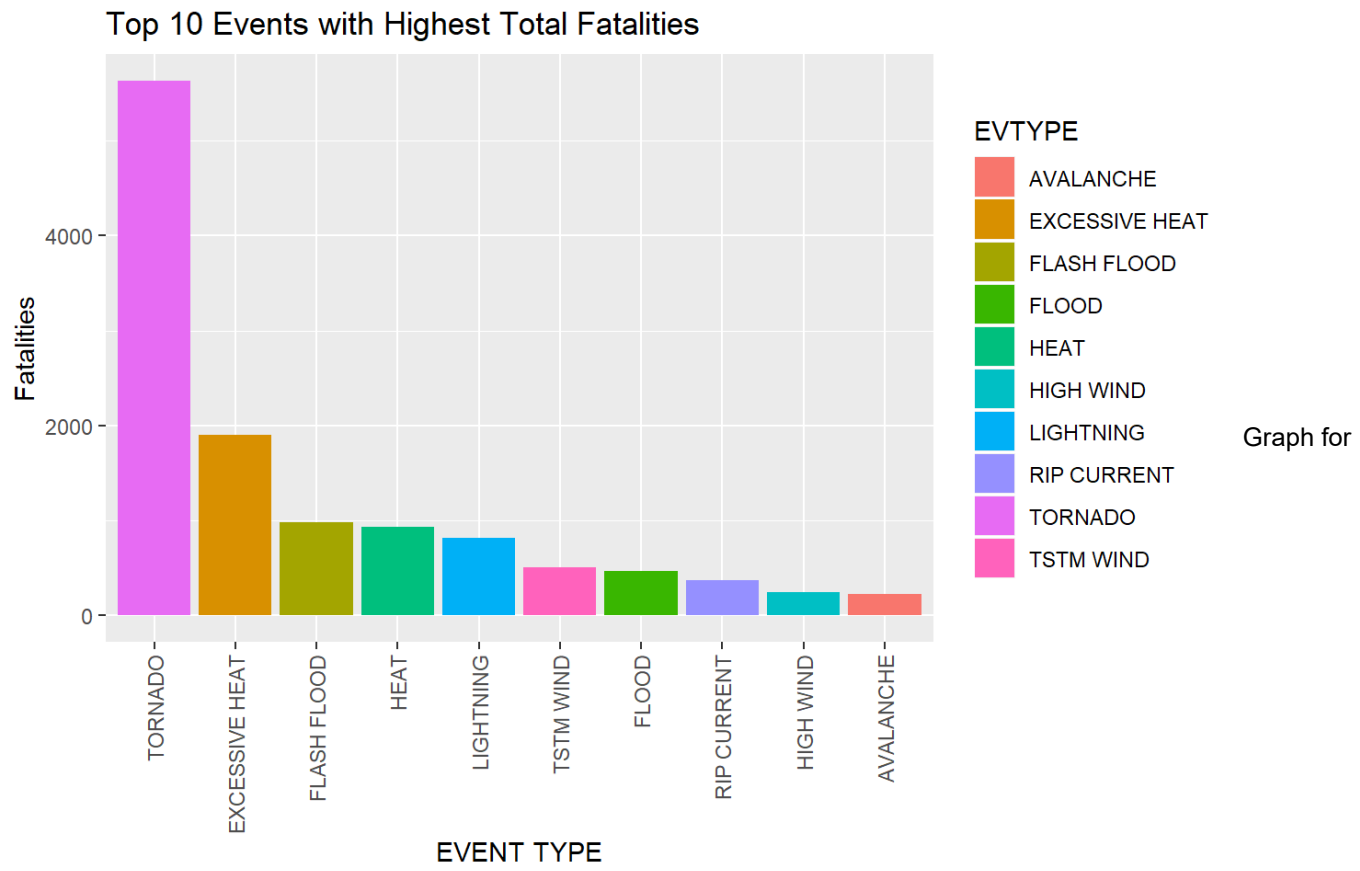
For the population health we will graph both the agregated top 10 of Fatalities and injuries by Event Type

Graph for the Top 10 Event with The Highest Fatalities indicate that Tornados has the most number of Fatalities and that tornados affect the most compared to other Event type.

```
library("ggplot2")
```

```
## Warning: package 'ggplot2' was built under R version 3.5.3
```

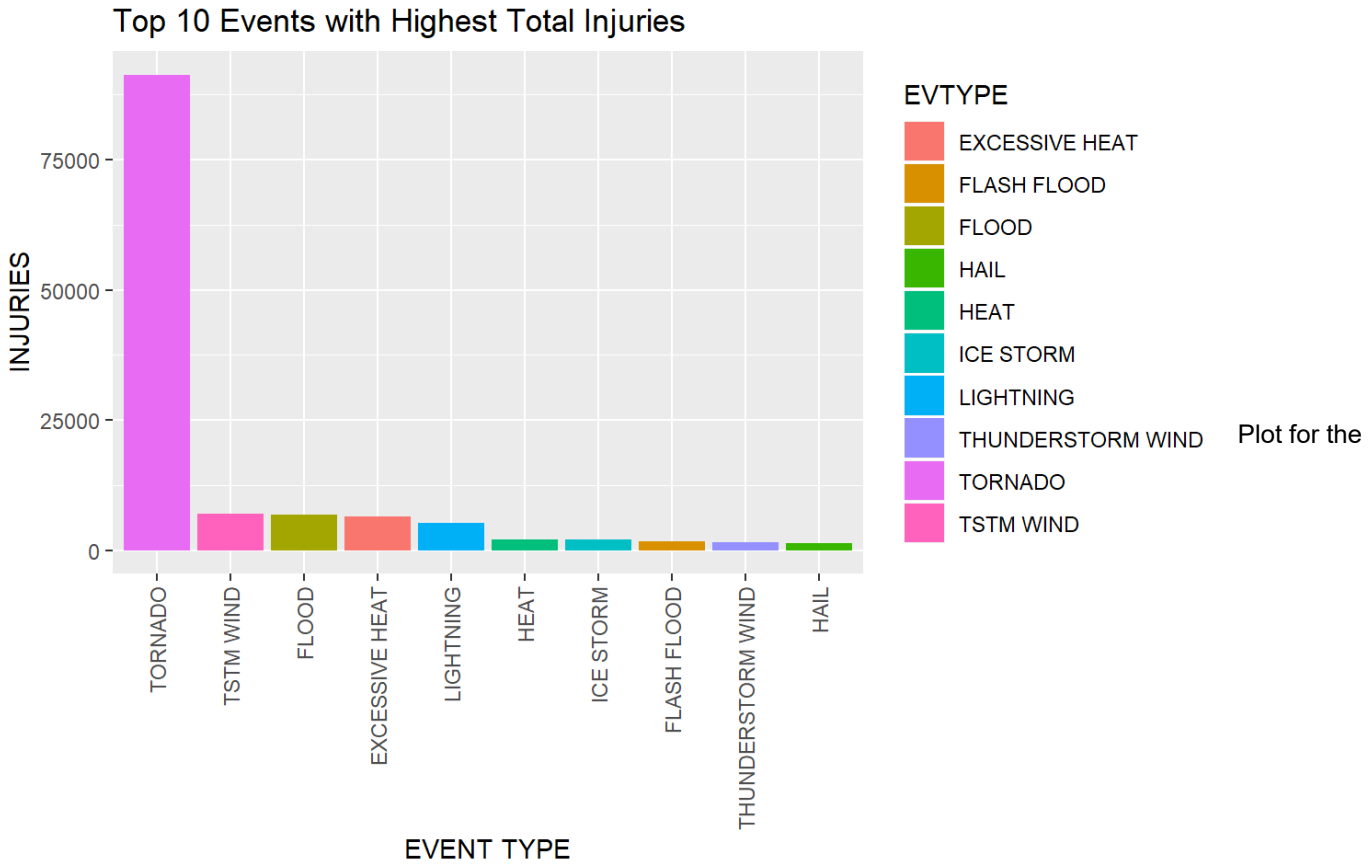
```
ggplot(data=StormData2.top10fatalities, aes(x=reorder(EVTYPE,-FATALITIES), y=FATALITIES, fill=EVTYPE))
+
        geom_bar(stat="identity") +
        ggtitle("Top 10 Events with Highest Total Fatalities") +
        labs(x="EVENT TYPE", y="Fatalities") +
        theme(axis.text.x = element_text(angle=90, vjust=0.5, hjust=1))
```



Graph for

the Top ten Events with the Highest Injuries indicate that Tornados has the most number of Injuries caused.It can also be seen how much the big difference in numbers when compared to other event Type

```
ggplot(data=StormData2.top10injuries, aes(x=reorder(EVTYPE,-INJURIES), y=INJURIES,fill=EVTYPE)) +
        geom_bar(stat="identity") +
        ggtitle("Top 10 Events with Highest Total Injuries") +
        labs(x="EVENT TYPE", y="INJURIES") +
        theme(axis.text.x = element_text(angle=90, vjust=0.5, hjust=1))
```

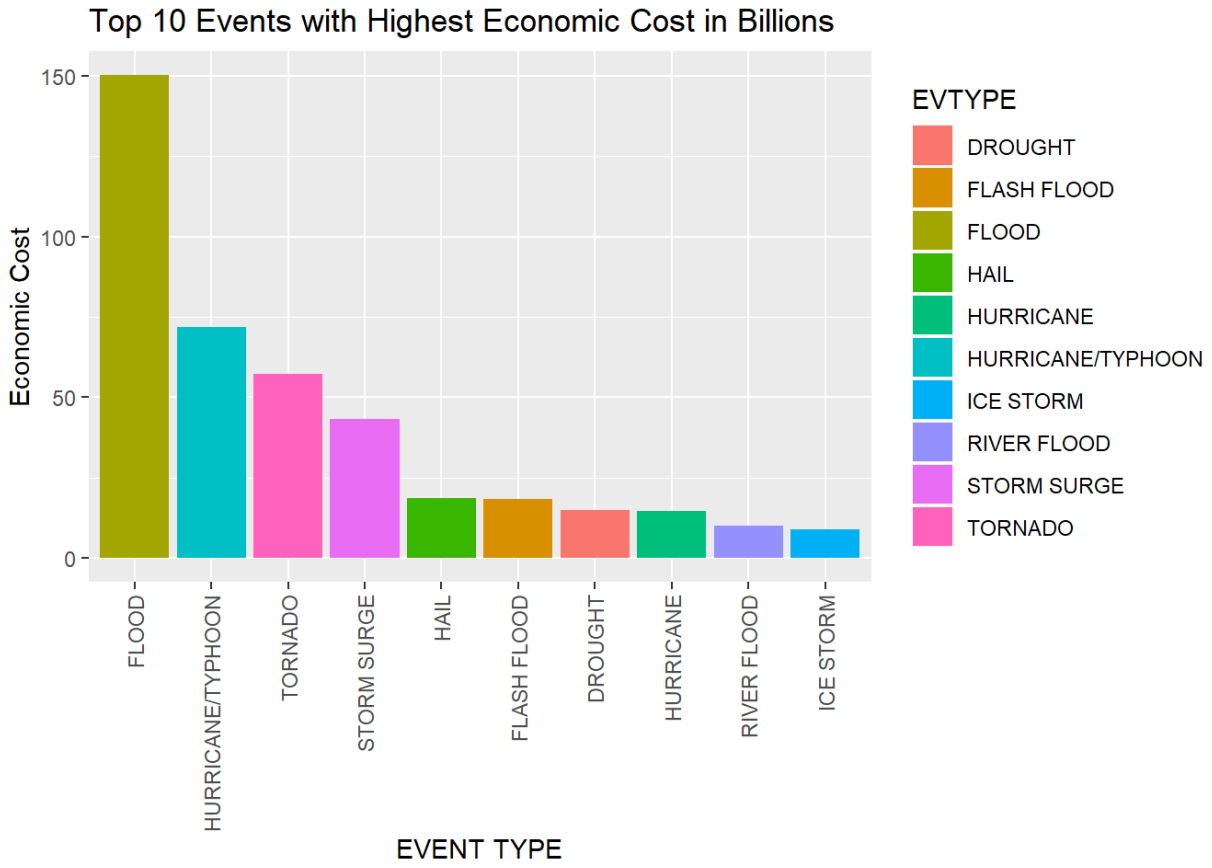


Plot for the

Economic Impact

The Top ten Events with Highest Economic Cost in Billions is from Hurricane that has the most Econimic Cost. It can also be seen how much the economic cost that the hurricanse caused.

```
ggplot(data=StormData2.top10eco, aes(x=reorder(EVTYPE,-ECODMG), y=ECODMG/1000000000,fill=EVTYPE)) +
        geom_bar(stat="identity") +
        ggtitle("Top 10 Events with Highest Economic Cost in Billions") +
        labs(x="EVENT TYPE", y="Economic Cost") +
        theme(axis.text.x = element_text(angle=90, vjust=0.5, hjust=1))
```



End of the Report