R Notebook to explore the NOAA Storm Database

Summary

The basic goal of this assignment is to explore the NOAA Storm Database and answer some basic questions about severe weather events.

The analysis addresses the following questions:

1. Across the United States, which types of events (as indicated in the EVTYPE variable) are most harmful with respect to population health?
2. Across the United States, which types of events have the greatest economic consequences?

We charge the libraries that we will use during the calculations and ploting

```
library(ggplot2)
library(dplyr) # to use case_when function
```

```
##
## 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
```

Data Processing

Fisrt we charge the data provided that can be found zipped in Data Storm

```
stormData <- read.csv("repdata_data_StormData.csv",na.strings="", sep=",", head=TRUE)
# show the first rows of data
head(stormData)
```

```
##   STATE__           BGN_DATE BGN_TIME TIME_ZONE COUNTY COUNTYNAME STATE
## 1       1  4/18/1950 0:00:00     0130       CST     97     MOBILE    AL
## 2       1  4/18/1950 0:00:00     0145       CST      3    BALDWIN    AL
## 3       1  2/20/1951 0:00:00     1600       CST     57    FAYETTE    AL
## 4       1   6/8/1951 0:00:00     0900       CST     89    MADISON    AL
## 5       1 11/15/1951 0:00:00     1500       CST     43    CULLMAN    AL
## 6       1 11/15/1951 0:00:00     2000       CST     77 LAUDERDALE    AL
##    EVTYPE BGN_RANGE BGN_AZI BGN_LOCATI END_DATE END_TIME COUNTY_END
## 1 TORNADO         0    <NA>       <NA>     <NA>     <NA>          0
## 2 TORNADO         0    <NA>       <NA>     <NA>     <NA>          0
## 3 TORNADO         0    <NA>       <NA>     <NA>     <NA>          0
## 4 TORNADO         0    <NA>       <NA>     <NA>     <NA>          0
## 5 TORNADO         0    <NA>       <NA>     <NA>     <NA>          0
## 6 TORNADO         0    <NA>       <NA>     <NA>     <NA>          0
##   COUNTYENDN END_RANGE END_AZI END_LOCATI LENGTH WIDTH F MAG FATALITIES
## 1         NA         0    <NA>       <NA>   14.0   100 3   0          0
## 2         NA         0    <NA>       <NA>    2.0   150 2   0          0
## 3         NA         0    <NA>       <NA>    0.1   123 2   0          0
## 4         NA         0    <NA>       <NA>    0.0   100 2   0          0
```

```
## 5         NA         0    <NA>       <NA>    0.0   150 2   0          0
## 6         NA         0    <NA>       <NA>    1.5   177 2   0          0
##   INJURIES PROPDMG PROPDMGEXP CROPDMG CROPDMGEXP  WFO STATEOFFIC ZONENAMES
## 1       15    25.0          K       0       <NA> <NA>       <NA>      <NA>
## 2        0     2.5          K       0       <NA> <NA>       <NA>      <NA>
## 3        2    25.0          K       0       <NA> <NA>       <NA>      <NA>
## 4        2     2.5          K       0       <NA> <NA>       <NA>      <NA>
## 5        2     2.5          K       0       <NA> <NA>       <NA>      <NA>
## 6        6     2.5          K       0       <NA> <NA>       <NA>      <NA>
##   LATITUDE LONGITUDE LATITUDE_E LONGITUDE_ REMARKS REFNUM
## 1     3040      8812       3051       8806    <NA>      1
## 2     3042      8755          0          0    <NA>      2
## 3     3340      8742          0          0    <NA>      3
## 4     3458      8626          0          0    <NA>      4
## 5     3412      8642          0          0    <NA>      5
## 6     3450      8748          0          0    <NA>      6
```

Explore the names of the columns to clasify the information.

```
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"
```

We see that for the firs question we are interested in the INJURIES and FATALITIES columns. For the second question we are interested in the "damage" columns (PROPDMG and CROPDMG).

First then we group the information by EVTYPE to have a reduced group adding the fatalities and injuries in two dataframes and order the values descending

```
eventFatalities <- aggregate(FATALITIES~EVTYPE, data=stormData, sum)
eventFatalities <- eventFatalities[order(-eventFatalities$FATALITIES),]

eventInjuries <- aggregate(INJURIES~EVTYPE, data=stormData, sum)
eventInjuries<-eventInjuries[order(-eventInjuries$INJURIES),]

eventFatalitiesAndInjuries <-aggregate((INJURIES +FATALITIES )~EVTYPE, data=stormData, sum)

eventFatalitiesAndInjuries<-eventFatalitiesAndInjuries[order(-eventFatalitiesAndInjuries$`(INJURIES + FA
```

We take only the ten greater values for fatalities and injuries to plot the values in the results section.

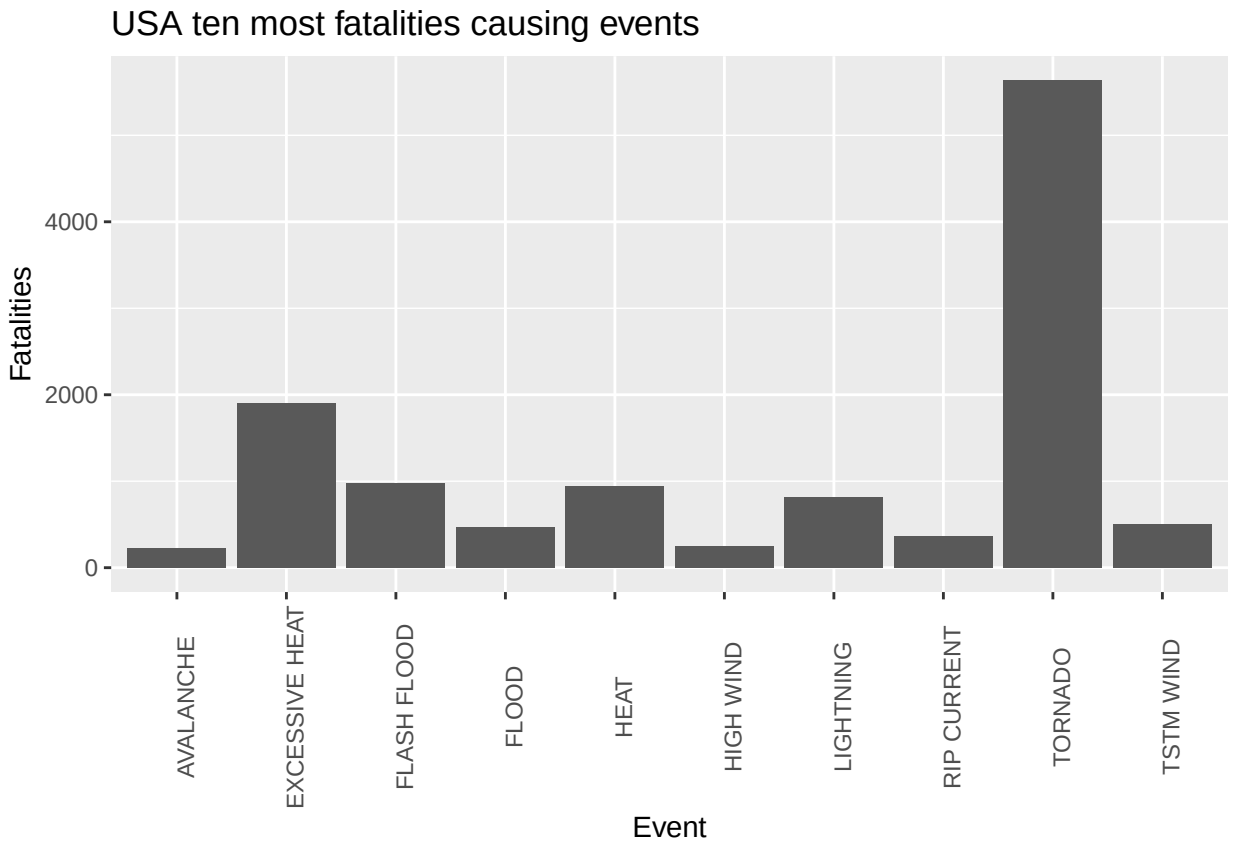
```
teneventFatalities <- head(eventFatalities, 10)
teneventInjuries <-head(eventInjuries, 10)
```

Results

From data procesing we see the most severe events, however a ploting is added to show it graphically.

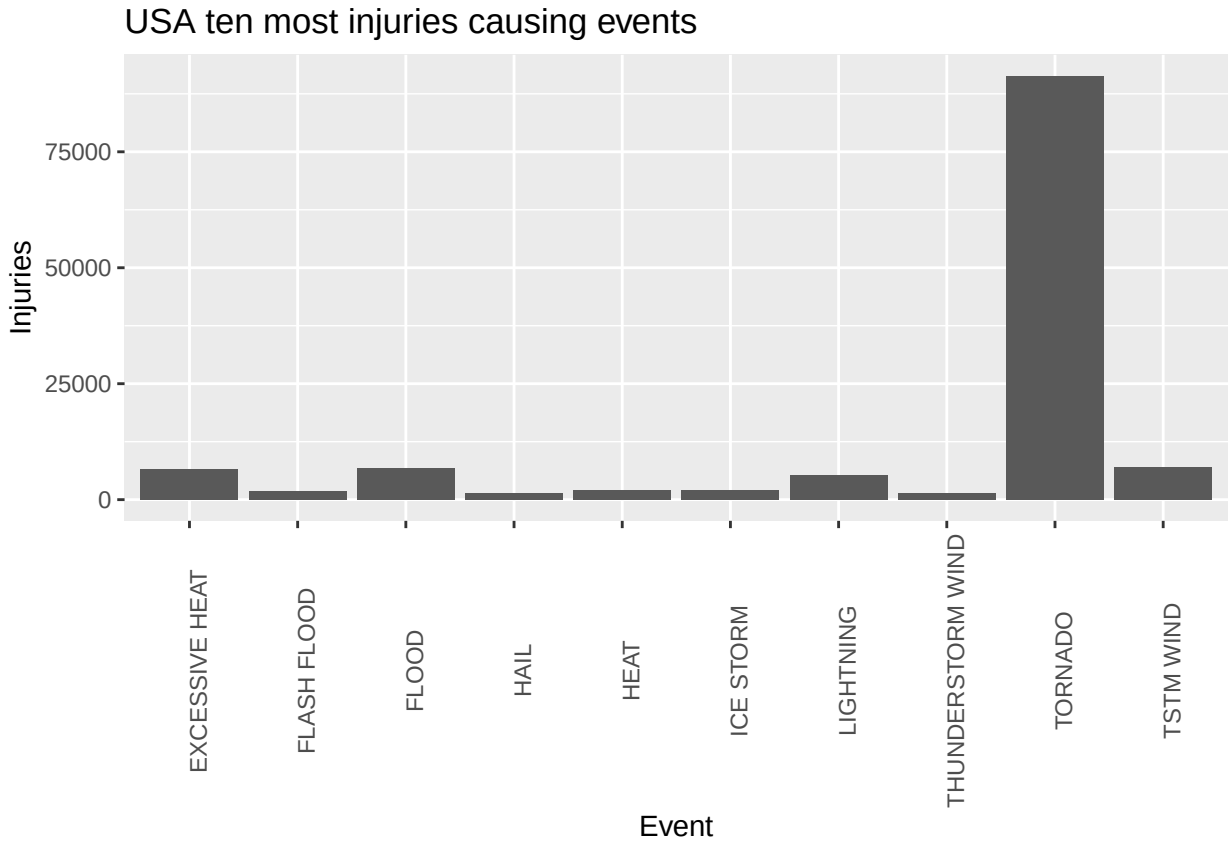
```
ggplot(data=teneventFatalities, aes(x=teneventFatalities$EVTYPE, y=teneventFatalities$FATALITIES)) +
    geom_bar(stat="identity",position=position_dodge()) +
```

```
    theme(axis.text.x = element_text(angle = 90)) +
    labs(x = "Event", y = "Fatalities", title = "USA ten most fatalities causing events")
```



The tornado is with great different the worst event followed by excesive heat and flash flood.

```
ggplot(data=teneventInjuries, aes(x=teneventInjuries$EVTYPE, y=teneventInjuries$INJURIES)) +
    geom_bar(stat="identity",position=position_dodge()) +
    theme(axis.text.x = element_text(angle = 90)) +
    labs(x = "Event", y = "Injuries", title = "USA ten most injuries causing events")
```



For injuries, the tornado, followed by TSTM Wind, Flood and excessive Heat. To keep the worst combined we just show the sum of both.

```
head(eventFatalitiesAndInjuries, 10)
```

```
##                EVTYPE (INJURIES + FATALITIES)
## 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
```

We consider then those values as the worst combining Injuries + fatalities.

For the second question we are going just to agregate all the damages and sort the worst events. Before doing the sum, we read in page 12 of the Storm Data Documentation And see that we have to convert to the same units before summing the values.

First we generate a copy of the dataframe and find different values to convert (from names we find the index of PROPDMGEXP and CROPDMGEXP)

```
eventsDamage <-stormData
unique(eventsDamage[,28] )
```

```
## [1] <NA> M    K    m    B    ?    0    k    2
```

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

```
unique(eventsDamage[,26] )
```

```
##  [1] K    M    <NA> B    m    +    0    5    6    ?    4    2    3    h
## [15] 7    H    -    1    8
## Levels: - ? + 0 1 2 3 4 5 6 7 8 B h H K m M
```

We generate a new column in the data set with the value in dolars of each damage to properties and crops and after that, we agregate the values in billions of dolars.

```
eventsDamage <-eventsDamage %>% mutate(damageprop = case_when(is.na(PROPDMGEXP) ~ PROPDMG,
                                            PROPDMGEXP=="K" ~ 1000*PROPDMG,
                                            PROPDMGEXP=="k" ~ 1000*PROPDMG,
                                            PROPDMGEXP=="M" ~ 1000000*PROPDMG,
                                            PROPDMGEXP=="m" ~ 1000000*PROPDMG,
                                            PROPDMGEXP=="B" ~ PROPDMG*1000000000,
                                            FALSE ~ PROPDMG))
eventsDamage <-eventsDamage %>% mutate(damagecrop = case_when(is.na(CROPDMGEXP) ~ CROPDMG,
                                            CROPDMGEXP=="K" ~ 1000*CROPDMG,
                                            CROPDMGEXP=="k" ~ 1000*CROPDMG,
                                            CROPDMGEXP=="M" ~ 1000000*CROPDMG,
                                            CROPDMGEXP=="m" ~ 1000000*CROPDMG,
                                            CROPDMGEXP=="B" ~ CROPDMG*1000000000,
                                            FALSE ~ CROPDMG))

#Agregate the damage in billions of Dolars
eventsDamage <-aggregate((damageprop+damagecrop)/1000000000~EVTYPE, data=eventsDamage, sum)
names(eventsDamage)[2]<- "damage"
eventsDamageOrdered<- eventsDamage[order(-eventsDamage$damage),]

head(eventsDamageOrdered, 10)
```

```
##                EVTYPE     damage
## 168             FLOOD 150.319678
## 407 HURRICANE/TYPHOON  71.913713
## 830           TORNADO  57.301936
## 666       STORM SURGE  43.323541
## 241              HAIL  18.733216
## 152       FLASH FLOOD  17.561539
## 94            DROUGHT  15.018672
## 398         HURRICANE  14.610229
## 586       RIVER FLOOD  10.148404
## 423         ICE STORM   8.967038
```

The table shows that the FOOD is the event generating the highest economical cost with more than 150 billion dolars.