Getting which Weather Event the most effects in population health and economy from NOAA Data

Getting which Weather Event the most effects in population health and economy

This analysis focuses in discovering which kind of weather event the most effects in population health and economy in the USA between 1950 and 2011. The data is recollected from the U.S. National Oceanic and Atmospheric Administration's NOAA service.

Data Processing

Obviously, the first step is preparing the environment. It means downloading the data and loading needed packages.

```
library(dplyr)
library(ggplot2)
library(gridExtra)

urlData <- "https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2"
pathData <- "data.csv.bz2"
if (!file.exists(pathData)) {
  download.file(urlData, destfile = "data.csv.bz2")
}
```

Later, it is only considered 14 variables based on the relation with the questions for the project and the available data in the data set (there exists a lot of NA values in those variables).

```
noaa <- read.csv(pathData, na.strings = "")
names_c <- c("BGN_DATE", "COUNTYNAME", "STATE", "EVTYPE", "LENGTH", "WIDTH", "F", "MAG", "FATALITIES",
noaa <- noaa[, names_c]
```

```
naPROP <- sum(is.na(noaa$PROPDMGEXP))
noaa$PROPDMGEXP <- ifelse(is.na(noaa$PROPDMGEXP), '0',noaa$PROPDMGEXP)
naCROP <- sum(is.na(noaa$CROPDMGEXP))
noaa$CROPDMGEXP <- ifelse(is.na(noaa$CROPDMGEXP), '0',noaa$CROPDMGEXP)
```

There are 465934 NA values in the variable property damage exponation.
There are 618413 NA values in the variable crop damage exponation.

So, it is neccesary fixing with 0 values since it will not generated any problems later.

Results

Question 1: Across the United States, which types of events (as indicated in the EVTYPE variable) are most harmful with respect to population health? For this question, The dplyer package is used to grouping based on type of event and summarise the total sum of fatalities and injuries. Additionally, observations with zero values in both variables (fatalities and injuries) are removed.

```
noaa_by_evtype <- noaa %>% group_by(EVTYPE) %>% summarise(tot_fatalities = sum(FATALITIES), tot_injuries
```

Over this new data set we order based on total injuries to see which event type causes the most damage in population health.
For injuries in population:

```
index_max_injuries <- which.max(noaa_by_evtype$tot_injuries)
event_injury <- noaa_by_evtype[index_max_injuries,1]
noaa_by_evtype[order(noaa_by_evtype$tot_injuries, decreasing = TRUE), c(1, 3)][1:10,]
```

```
## # A tibble: 10 x 2
##    EVTYPE            tot_injuries
##    <fct>                    <dbl>
##  1 TORNADO                  91346
##  2 TSTM WIND                 6957
##  3 FLOOD                     6789
##  4 EXCESSIVE HEAT            6525
##  5 LIGHTNING                 5230
##  6 HEAT                      2100
##  7 ICE STORM                 1975
##  8 FLASH FLOOD               1777
##  9 THUNDERSTORM WIND         1488
## 10 HAIL                      1361
```

It shows clearly that 834 is the weather event which generates the most damage causing injuries.
Similarly, for fatalities in population:

```
index_max_fatalities <- which.max(noaa_by_evtype$tot_fatalities)
event_fatality <- noaa_by_evtype[index_max_fatalities,]
noaa_by_evtype[order(noaa_by_evtype$tot_fatalities, decreasing = TRUE), c(1, 2)][1:10,]
```

```
## # A tibble: 10 x 2
##    EVTYPE         tot_fatalities
##    <fct>                   <dbl>
##  1 TORNADO                  5633
##  2 EXCESSIVE HEAT           1903
##  3 FLASH FLOOD               978
##  4 HEAT                      937
##  5 LIGHTNING                 816
##  6 TSTM WIND                 504
##  7 FLOOD                     470
##  8 RIP CURRENT               368
##  9 HIGH WIND                 248
## 10 AVALANCHE                 224
```

It shows clearly that 834, 5633, 91346 is the weather event which generates the most damage causing fatalities.

So based on evidences 834, 5633, 91346 generates the most health damage between wheather events.

Question 2: Across the United States, which types of events have the greatest economic consequences? For the second question we need to clean a little more the data. Basically, we have to convert all the economic information to the same unit. For that reason, *convertWIthBase* function helps to convert all units to the same format.

```
convertWithBase <- function(v) {
```

```
  val <- as.numeric(v[1])
  b <- v[2]
  c <- substr(b, 1, 1)
  numb <- c("0","1","2","3","4","5","6","7","8","9")
  if (c == '-' || c == '?' || c == '+') {
    val
  } else if (c %in% numb) {
    val * 10 ** as.numeric(b)
  } else if (c == 'h' || c == 'H') {
    val * 100
  } else if (c == 'k' || c == 'K') {
    val * 1000
  } else if (c == 'm' || c == 'M') {
    val * 1000000
  } else if (c == 'b' || c == 'B') {
    val * 1000000000
  } else {
    val
  }
}

noaa$CROPDMGTOT <- apply(data.frame(crop=noaa$CROPDMG, exp=noaa$CROPDMGEXP), MARGIN = 1, FUN = convertW

noaa$PROPDMGTOT <- apply(data.frame(crop=noaa$PROPDMG, exp=noaa$PROPDMGEXP), MARGIN = 1, FUN = convertW
```

Later, with all data with the same economic unit, it is necessary summarize by damage in properties and crops. With the intention to get more visibility I applied a filter to get the events with the most economic effect.

```
noaa_by_evtype_econ <- noaa %>% group_by(EVTYPE) %>% summarise(tot_prop = sum(PROPDMGTOT), tot_crop = su
```

One first table show the most 5 weather events which causing the most damage in properties.

```
event_prop <- noaa_by_evtype_econ[which.max(noaa_by_evtype_econ$tot_prop),1]
noaa_prop <- noaa_by_evtype_econ[order(noaa_by_evtype_econ$tot_prop, decreasing = T),c(1, 2)][1:5,]
noaa_prop
```

```
## # A tibble: 5 x 2
##   EVTYPE      tot_prop
##   <fct>          <dbl>
## 1 TORNADO      8.01e22
## 2 FLOOD        3.01e22
## 3 FLASH FLOOD  2.78e22
## 4 HAIL         2.00e22
## 5 TSTM WIND    1.65e22
```

In similar way, the table show the most 5 weather events which causing the most damage in crops.

```
event_crop <-noaa_by_evtype_econ[which.max(noaa_by_evtype_econ$tot_crop), 1]
noaa_crop <- noaa_by_evtype_econ[order(noaa_by_evtype_econ$tot_crop, decreasing = T), c(1, 3)][1:7,]
noaa_crop
```

```
## # A tibble: 7 x 2
##   EVTYPE            tot_crop
##   <fct>                <dbl>
## 1 DROUGHT      1266558015000
```

```
## 2 HAIL          821632152003
## 3 FLOOD         712481450000
## 4 FLASH FLOOD   302293100000
## 5 HURRICANE     276531000000
## 6 TSTM WIND     153282350000
## 7 EXTREME COLD  133647000000
```

Finally, the table show the most 5 weather events which causing the most damage in crops and properties in total.

```
noaa_by_evtype_econ$tot_dam <- noaa_by_evtype_econ$tot_prop + noaa_by_evtype_econ$tot_crop
noaa_dam <- noaa_by_evtype_econ[order(noaa_by_evtype_econ$tot_dam, decreasing = T), c(1, 4)][1:6,]
noaa_dam
```

```
## # A tibble: 6 x 2
##   EVTYPE            tot_dam
##   <fct>               <dbl>
## 1 TORNADO           8.01e22
## 2 FLOOD             3.01e22
## 3 FLASH FLOOD       2.78e22
## 4 HAIL              2.00e22
## 5 TSTM WIND         1.65e22
## 6 THUNDERSTORM WIND 1.13e22
```

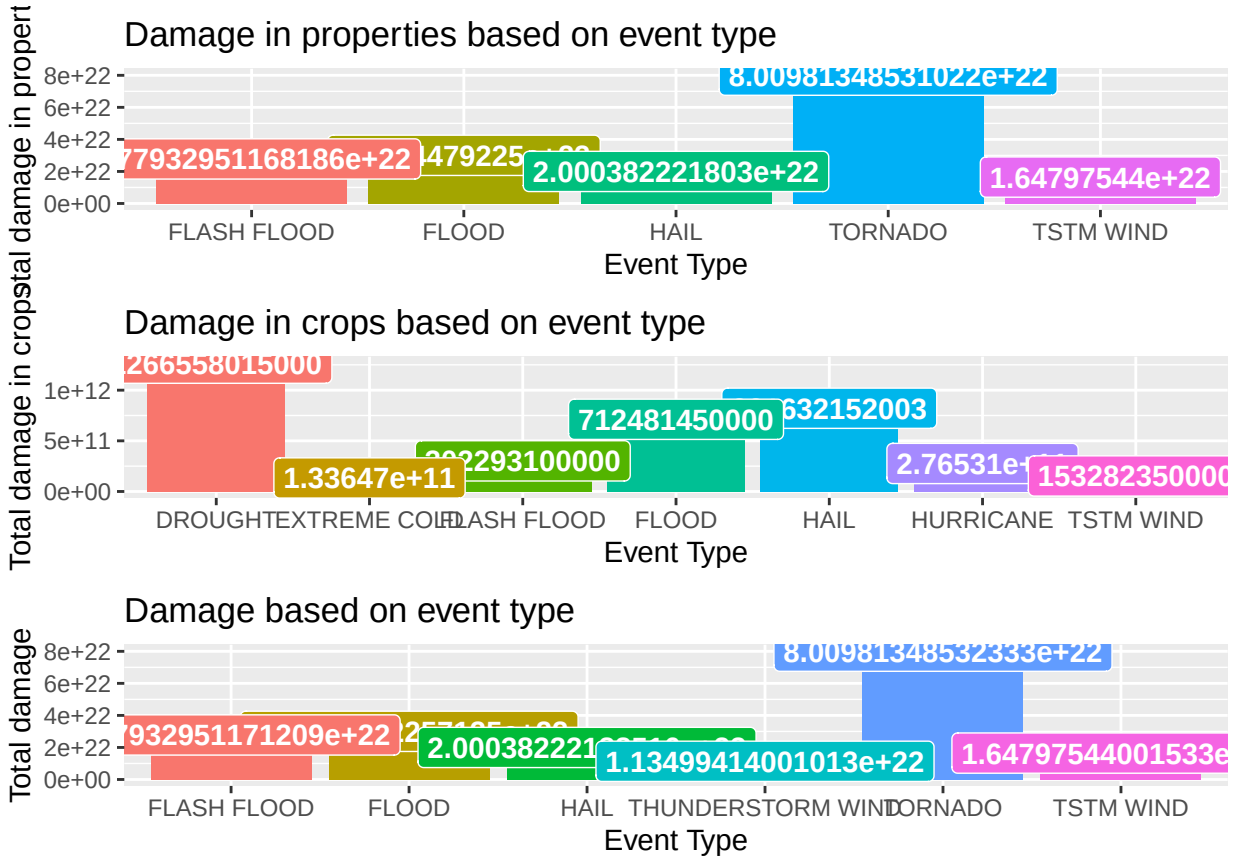
For a better visualization, one plot with the 3 criterios is showed.

```
plot_prop <- ggplot(noaa_prop, aes(x=EVTYPE, y=tot_prop,fill=EVTYPE, label = round(tot_prop,2))) +
    geom_bar(stat="identity") +
    xlab("Event Type") +
    ylab("Total damage in properties") +
    ggtitle("Damage in properties based on event type")+
    geom_label(aes(fill = EVTYPE),colour = "white", fontface = "bold") +
    theme(legend.position="none")

plot_crop <- ggplot(noaa_crop, aes(x=EVTYPE, y=tot_crop,fill=EVTYPE, label = round(tot_crop,2))) +
    geom_bar(stat="identity") +
    xlab("Event Type") +
    ylab("Total damage in crops") +
    ggtitle("Damage in crops based on event type")+
    geom_label(aes(fill = EVTYPE),colour = "white", fontface = "bold") +
    theme(legend.position="none")

plot_dam <- ggplot(noaa_dam, aes(x=EVTYPE, y=tot_dam,fill=EVTYPE, label = round(tot_dam,2))) +
    geom_bar(stat="identity") +
    xlab("Event Type") +
    ylab("Total damage") +
    ggtitle("Damage based on event type")+
    geom_label(aes(fill = EVTYPE),colour = "white", fontface = "bold") +
    theme(legend.position="none")

fig <- function(width, heigth){
     options(repr.plot.width = width, repr.plot.height = heigth)
}
fig(1000, 4000)
options(repr.plot.width = 14, repr.plot.height = 8)
grid.arrange(plot_prop, plot_crop, plot_dam, nrow=3)
```



For this case, it shows that *TORNADO* is the weather event which generates the most damage in economic.