W4Assignment

7cats

30/07/2020

TITLE: Analysis of Damage Cost by Weather Phenomenon based NOAA Storm data

Processing Methods: First events without actual damage are removed for efficiency, then types of events are categorized into 15 classes manually. Damages are calculated in two ways, the sum of each type of event and the average damage costed by one single events within each type. While calculating the damage, the unit of recorded damage should be considered.

ANALYSIS: The data is stored in file "repdata_data_StormData.csv.bz2" and read into variable metaData

```
metaData <- read.csv("repdata_data_StormData.csv.bz2")
```

First events with zero property/crop damage and fatalities/injuries are removed from the table for simplicity

```
metaData <- metaData %>%
    mutate(damage = PROPDMG + FATALITIES + INJURIES + CROPDMG) %>%
    .[.$damage != 0,]
```

Analyse the relationship between event type and population health

```
keyType <- paste( c("FLOOD","TEMP.","WIND","ICE","STORM","RAIN","FIRE","FOG","RIP CURRENT","SLIDE","DRY
metaData <- metaData %>%
    mutate(EVTYPE = toupper(EVTYPE),
           EVTYPE = str_replace_all(EVTYPE,"WAYTERSPOUT|WATER SPOUT","WATERSPOUT"),
           EVTYPE = str_replace_all(EVTYPE,"TSTM|THUNDERSTORM|TORNADO|TYPHOON|HURRICANE|TROPICAL STORM|
                                    |TROPICAL DEPRESSION|WATERSPOUT|WALL CLOUD|GUSTNADO|BURST|TURBULENC
           EVTYPE = str_replace_all(EVTYPE,"WINDCHILL|WINTER WEATHER|COOL|WINTER|LOW TEMP|WINTRY|HOT|WA
                                    |HIGH TEMPERATURE|TEMPERATURE|COLD|RECORD","TEMP."),
           EVTYPE = str_replace_all(EVTYPE,"FLD|URBAN|DAM|STREAM|FLOOODING","FLOOD"),
           EVTYPE = str_replace_all(EVTYPE,"DROUGHT|DRIEST","DRY"),
           EVTYPE = str_replace_all(EVTYPE,"PRECIPITATION|PRECIP|SHOWERS|SHOWER|DRIZZLE","RAIN"),
           EVTYPE = str_replace_all(EVTYPE,"AVALANCHE|SLUMP|EROSION|EROSIN|AVALANCE","SLIDE"),
           EVTYPE = str_replace_all(EVTYPE,"VOLCAN|SMOKE","DUST"),
           EVTYPE = str_replace_all(EVTYPE,"MARINE MISHAP|APACHE COUNTY|MARINE ACCIDENT|HEAVY MIX|DROWN
           EVTYPE = str_replace_all(EVTYPE,"WND","WIND"),
           EVTYPE = str_replace_all(EVTYPE,"LIGHTING|LIGHTNING|LIGHTS|LIGNTNING","LIGHT"),
           EVTYPE = str_replace_all(EVTYPE,"HAIL|WINTER STORM|SLEET|SNOW|BLIZZARD|FREEZE|FROST|ICY|GLAZ
           EVTYPE = str_replace_all(EVTYPE,"SURF|TIDE|STORM SURGE|HIGH WAVES|HIGH WATER|COASTAL SURGE|TS
           |SWELLS|SEICHE|HIGH SEAS|RISING WATER|HEAVY SEAS|ROUGH SEAS|ROGUE WAVE","HWL"),
           EVTYPE = ifelse(is.na(str_extract(EVTYPE,keyType)),EVTYPE,str_extract(EVTYPE,keyType))
    )
unique(metaData$EVTYPE)
```

```
##  [1] "STORM"       "ICE"         "TEMP."       "RAIN"        "LIGHT"
##  [6] "FOG"         "RIP CURRENT" "FLOOD"       "OTHER"       "WIND"
## [11] "SLIDE"       "DUST"        "HWL"         "FIRE"        "DRY"
## [16] "WET"
```

All weather phenomenons are reduced to 15 types: Flood, Temp., Wind, Ice, Storm, Rain, Fire, Fog, Rip Current, Slide, Dry, Wet, Dust, other, High Water Level, Light. The classification might not be accurate but it's simply used as a method in this assignment. After classified, the data will be grouped and each group's relation with population damage will be analysed. First to calculate the economic damage the unit of damage and value should be combined. The unit of damage is recorded in variable: "PROPDMGEXP" and "CROPDMGEXP". Some rows contain abnormal unit are excluded from the dataset.

```
metaData <- metaData %>%
    mutate(CROPDMGEXP = toupper(CROPDMGEXP)) %>%
    mutate(PROPDMGEXP = toupper(PROPDMGEXP)) %>%
    filter(grepl(paste( c("K","B","H","M"), collapse = "|"), CROPDMGEXP)) %>%
    filter(grepl(paste( c("K","B","H","M"), collapse = "|"), PROPDMGEXP)) %>%
    mutate(unitCrop = sapply(CROPDMGEXP, function(x)
        switch(grep(x,c("H","K","M","B")), 100, 1000, 1e+06, 1e+09)),
           unitProp = sapply(PROPDMGEXP, function(x)
        switch(grep(x,c("H","K","M","B")), 100, 1000, 1e+06, 1e+09))) %>%
    mutate(CROPDMG = CROPDMG*unitCrop, PROPDMG = PROPDMG*unitProp)
```

The damage to population health is measured with fatalities and injuries separately and the economic consequences are computed by crop damage and property damage. First both categories and summed.

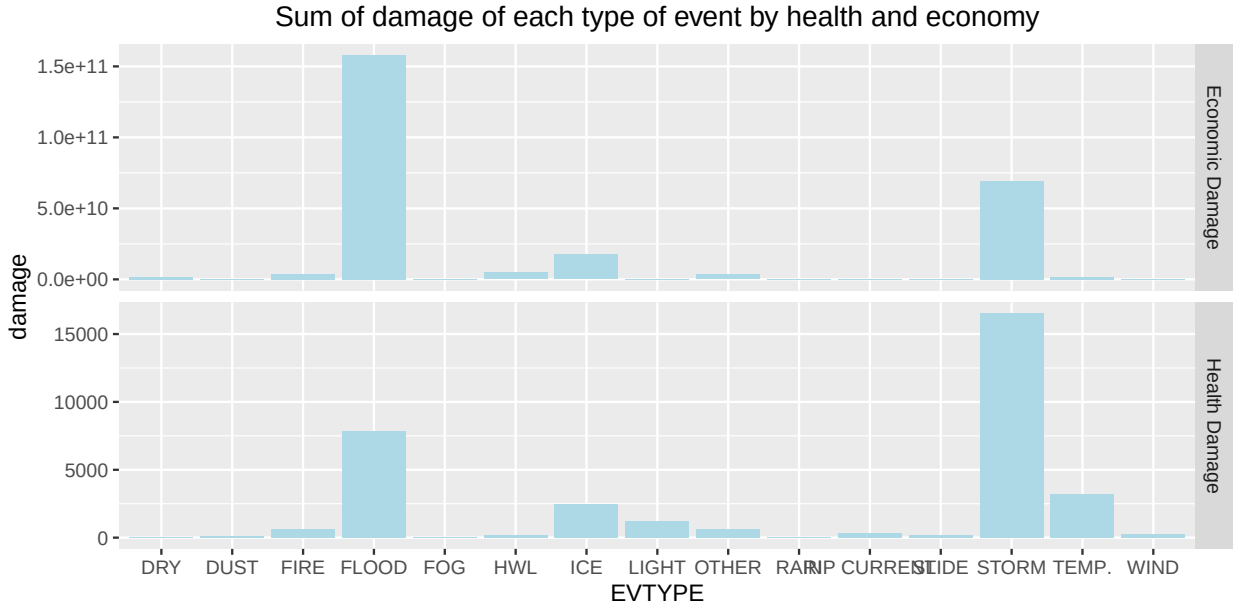
```
sumData <- metaData %>%
    group_by(EVTYPE) %>%
    summarise(
        healDamg = sum(FATALITIES, na.rm = T) + sum(INJURIES, na.rm = T),
        econDamg = sum(CROPDMG, na.rm = T) +sum(PROPDMG, na.rm = T)
        )
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
sumData <- sumData %>%
    gather(key = "damageType", value = "damage", healDamg:econDamg)
```

```
damgLabel <- c("Economic Damage", "Health Damage")
names(damgLabel) <- c("econDamg", "healDamg")

theme_update(plot.title = element_text(hjust = 0.5))
g <- ggplot(sumData, aes(x = EVTYPE, y = damage))
g + facet_grid(rows = vars(damageType), scales="free",
               labeller = labeller(damageType = damgLabel)) +
    geom_bar(stat="identity", fill = "light blue") +
    ggtitle("Sum of damage of each type of event by health and economy")
```




```
dev.copy(png,"SumDamageComparison.png",width = 960, height = 480)
```

```
## png
##   3
```

As it is shown in the figure, the type of event that has the highest health damage is flooding, and the storm type of event costs the most on economy. However this is just the sum of all events within one type, while the mean of all events is also worth investigating. Therefore the mean of each type of event is analyzed in the following.

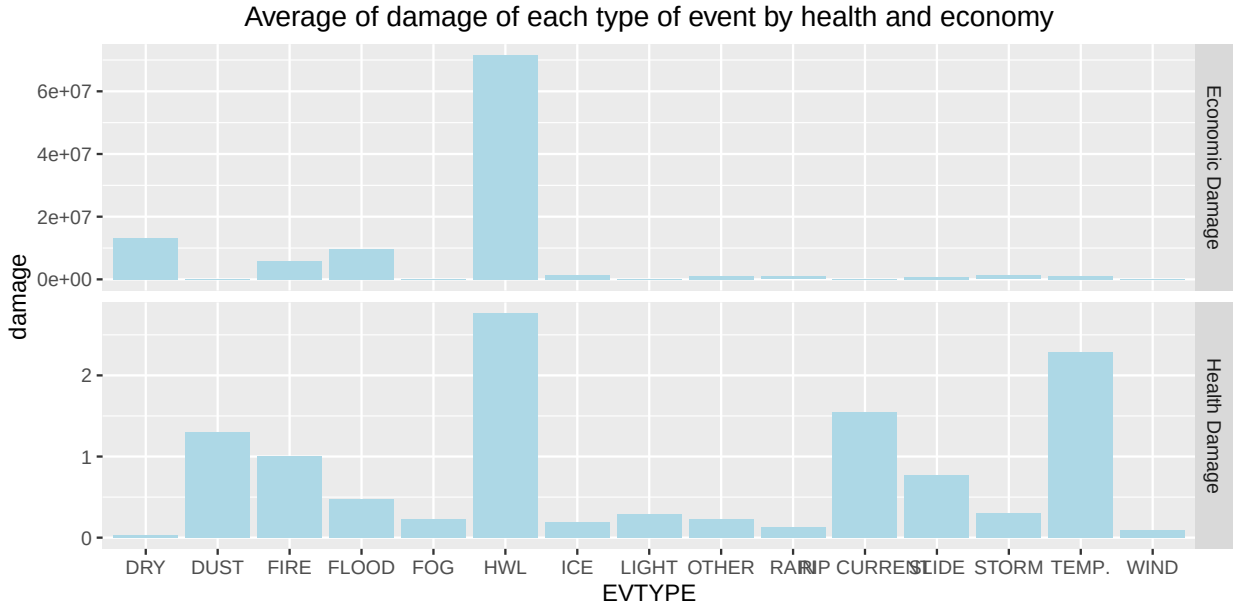
```
aveData <- metaData %>%
    group_by(EVTYPE) %>%
    summarise(
        healDamg = mean(FATALITIES, na.rm = T) + mean(INJURIES, na.rm = T),
        econDamg = mean(CROPDMG, na.rm = T) + mean(PROPDMG, na.rm = T)
        )
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
aveData <- aveData %>%
    gather(key = "damageType", value = "damage", healDamg:econDamg)
```

And the result of the damage averaged by each event is present in the following.

```
g <- ggplot(aveData, aes(x = EVTYPE, y = damage))
g + facet_grid(rows = vars(damageType), scales="free",
               labeller = labeller(damageType = damgLabel)) +
    geom_bar(stat="identity", fill = "light blue") +
    ggtitle("Average of damage of each type of event by health and economy")
```




```
dev.copy(png,"AveDamageComparison.png",width = 960, height = 480)
```

```
## png
##   4
```

As it is shown in the figure, the High Water Level damage has the most damage on both population health and economy. This suggest although flooding and storm happen more frequently and cost more in total, the damage resulted by one single High Water Level event is far higher.