Storm data assessment

Ania

15 12 2021

##Analysis of weather events severity in United States between 1950 and 2011

The file presents analysis of Storm Data collected between 1950 and 2011 by U.S. National Oceanic and Atmospheric Administration's (NOAA).The data set contains information about different weather connected events, their duration, fatalities or injured people as well as property damage.

The file describes the whole process of data analysis from data downloading the to the final results. In the analysis I indicated the most harmful (injuries and fatalities) types of weather events and analysed what kind of events have the greatest economic consequences.

The analysis was performed using three packages: *downloader* (data downloading), *dplyr*(data processing), *ggplot* and *cowplot* (graph plotting). The analysis is divided in three subsection: - Data Pre-processing, - Analysis of the impact of weather conditions on human life (injuries and fatalities), - Analysis of economic impact of weather conditions

Data Pre-processing

Data were loaded from the U.S. National Oceanic and Atmospheric Administration's website. CSV file was read into RStudio and data containing information about injuries, fatalities, crop and property damage as well as event type were subtracted to the new data frame.

```
#loading required libraries

library(downloader)
```

```
## Warning: pakiet 'downloader' został zbudowany w wersji R 4.1.1
```

```
library(ggplot2)
library(cowplot)
```

```
## Warning: pakiet 'cowplot' został zbudowany w wersji R 4.1.2
```

```
library(dplyr)
```

```
## Warning: pakiet 'dplyr' został zbudowany w wersji R 4.1.1
```

```
##
## Dołączanie pakietu: 'dplyr'
```

```
## Następujące obiekty zostały zakryte z 'package:stats':
##
##     filter, lag

## Następujące obiekty zostały zakryte z 'package:base':
##
##     intersect, setdiff, setequal, union
```

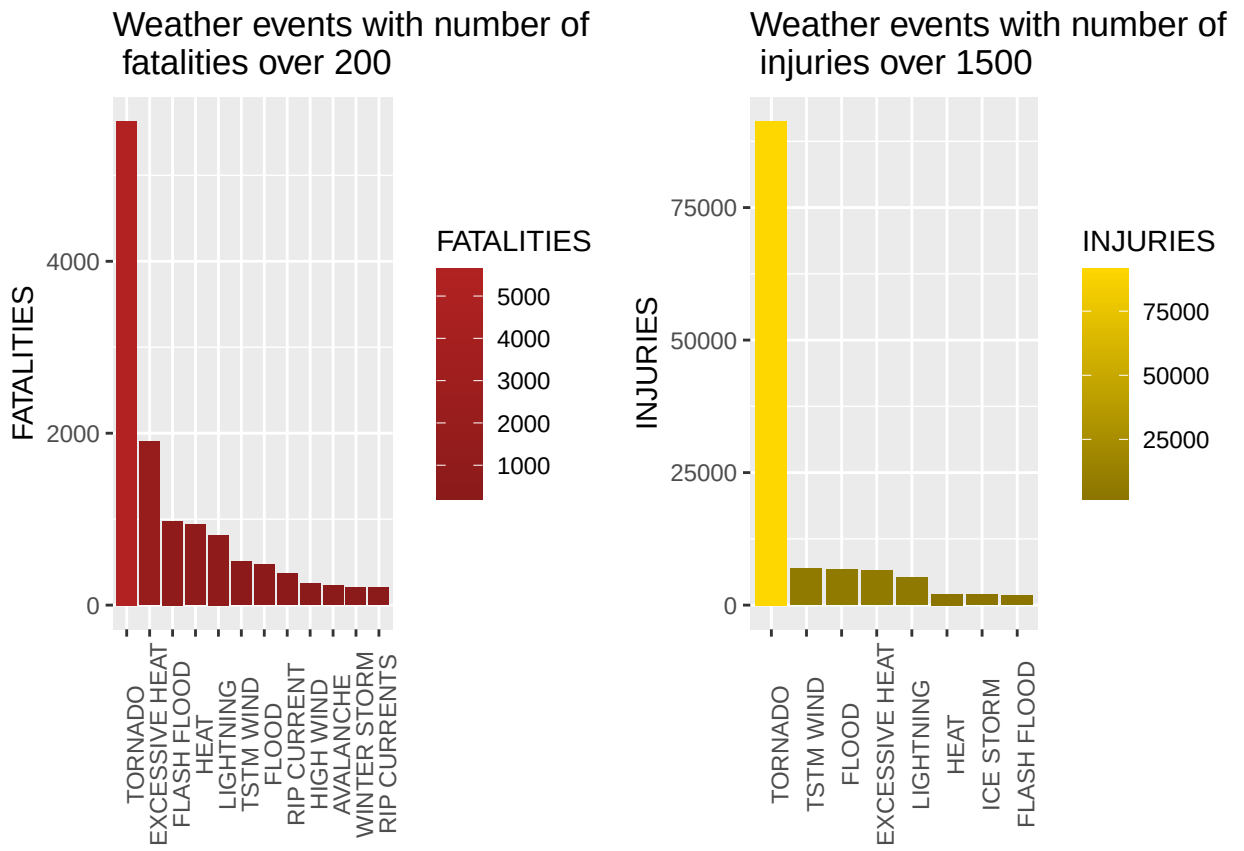
```
#downloading data file

filename <- "Storm data.csv"

dataURL <- "https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2"

if(!file.exists(filename)) {
  download(dataURL, filename, mode = "wb")
}

#loading data set

Data <- read.csv("Storm data.csv", header = TRUE)
StormData <- Data[,c("EVTYPE", "FATALITIES","INJURIES", "PROPDMG", "PROPDMGEXP","CROPDMG", "CROPDMGEXP"]
#StormData$PROPDMGEXP <- as.factor(StormData$PROPDMGEXP)
#StormData$CROPDMGEXP <- as.factor(StormData$CROPDMGEXP)
```

Analysis of the impact of weather conditions on human life (injuries and fatalities)

To show the number of fatalities and injured people for each event the **aggregate** function was used and the number of fatalities/injuries for the same events were summed up. From the data set events with fatalities number higher than 200 and events with injuries over 1500 were chosen for further analysis. For each outcome the separate bar plots were created and merged in one plot panel with *plot_grid* function. Panel plot was saved in wd with *dev.copy* function.

```
#calculating the number of fatalities and injuries for each event type

FatalInjuriesData <- aggregate(cbind(FATALITIES,INJURIES)~EVTYPE, StormData, FUN = sum)

#subtracting events with number of fatalities higher than 200 and injuties higher than 1500

FatalData <- FatalInjuriesData[FatalInjuriesData$FATALITIES>200,]
InjuriesData <- FatalInjuriesData[FatalInjuriesData$INJURIES>1500,]

#plotting the results

p1 <- ggplot(data =FatalData, aes(x = reorder(EVTYPE, -FATALITIES), y = FATALITIES))

p1 <- p1 + geom_bar(stat = "identity", aes(fill = FATALITIES)) + labs(y = "FATALITIES", x = NULL) + ther

p2 <- ggplot(data = InjuriesData, aes(x = reorder(EVTYPE, -INJURIES), y = INJURIES))

p2 <- p2 + geom_bar(stat = "identity", aes(fill = INJURIES)) + labs(y = "INJURIES", x = NULL) + theme(a
```

```
#creating the panel plot and saving the results

plot_grid(p1,p2)
```




```
dev.copy(png, "plotpanel1.png", width = 780, height = 480)
```

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

```
dev.off()
```

```
## pdf
##   2
```

Analysis showed that tornadoes have the highest impact on human life of all analysed events types. The number of injured people was much higher than for other analysed events.

3. Analysis of economic impact of weather conditions

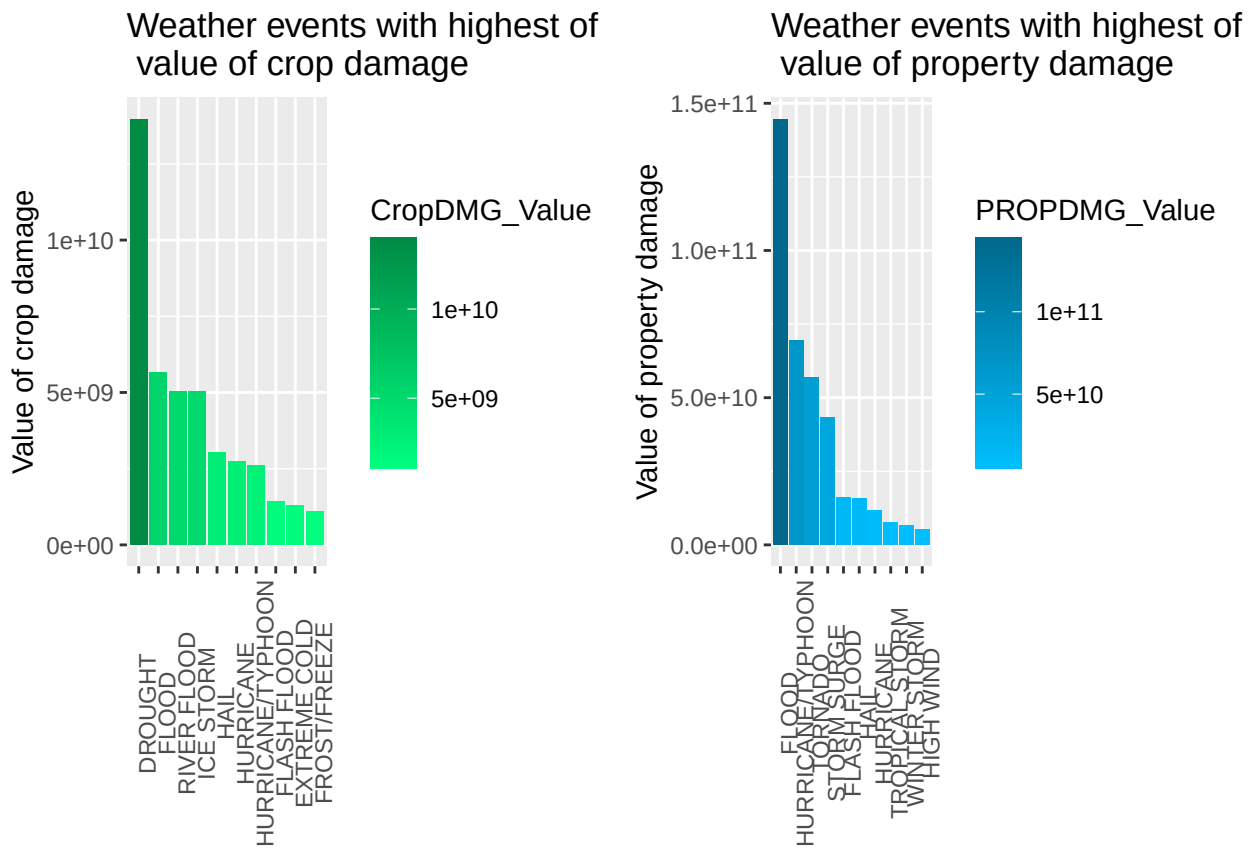
To calculate the value of damage caused by different types of events in the first step of data analysis function *multi_exp* was created. The function changes the letter expressions of the exponents of the powers to numerical values, which in further stages of the analysis will be used to calculate the final damage values.

With *sapply* function final value of damage is calculated and sign in new columns both for crop and property damage.

In the next step on analysis sum of total value of damage for crops and properties caused by each event was calculated with function *aggregate*. Top 10 values for crop and properties damage were subtracted to the new data frames with *top_n* function from *dplyr* package. For each of the damage types the separate bar plots were created and merged in one plot panel with *plot_grid* function. Panel plot was saved in wd with *dev.copy* function.

```
EconomicDMG <- StormData[StormData$PROPDMG>0 | StormData$CROPDMG>0,]
EconomicDMG$PROPDMGEXP<- as.factor(EconomicDMG$PROPDMGEXP)
EconomicDMG$CROPDMGEXP<- as.factor(EconomicDMG$CROPDMGEXP)

#creating the function for changing letters letter expressions of the exponents of the powers to numeri

multi_exp <- function(e) {

    if (e %in% c("h", "H"))
    return(2)

    else if (e %in% c("k", "K"))
    return(3)

    else if (e %in% c("m", "M"))
    return(6)

    else if (e %in% c("b", "B"))
    return(9)

    else {
      return(0)
    }
}

# calculating the final value of damages and saving the results in new columns

EconomicDMG$CROPDMGEXP_numeric <- sapply(EconomicDMG$CROPDMGEXP, FUN = multi_exp)
EconomicDMG$CropDMG_Value <- EconomicDMG$CROPDMG*(10**EconomicDMG$CROPDMGEXP_numeric)

EconomicDMG$PROPDMGEXP_numeric <- sapply(EconomicDMG$PROPDMGEXP, FUN = multi_exp)
EconomicDMG$PROPDMG_Value <- EconomicDMG$PROPDMG*(10**EconomicDMG$PROPDMGEXP_numeric)

#calculating the value of damage for each event and subtracting 10 top values for crop and properties d

CropDamageValue <- aggregate(CropDMG_Value~EVTYPE, EconomicDMG, FUN = sum)
CropDamageValue_top10 <- top_n(CropDamageValue[(order(CropDamageValue$CropDMG_Value, na.last = TRUE,deci


## Selecting by CropDMG_Value

PropDamageValue <- aggregate(PROPDMG_Value~EVTYPE, EconomicDMG, FUN = sum)
PropDamageValue_top10 <- top_n(PropDamageValue[(order(PropDamageValue$PROPDMG_Value, na.last = TRUE,deci


## Selecting by PROPDMG_Value
```

```
#plotting the results

p3 <- ggplot(CropDamageValue_top10, aes(x = reorder(EVTYPE, - CropDMG_Value), y = CropDMG_Value))

p3 <- p3 + geom_bar(stat = "identity", aes(fill = CropDMG_Value)) + labs(y = "Value of crop damage", x =

p4 <- ggplot(PropDamageValue_top10, aes(x = reorder(EVTYPE, - PROPDMG_Value), y = PROPDMG_Value))

p4 <- p4 + geom_bar(stat = "identity", aes(fill = PROPDMG_Value)) + labs(y = "Value of property damage"

#creating the panel plot and saving the results

plot_grid(p3,p4)
```




```
dev.copy (png, "panelplot2.png", width=780, height=480)
```

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

```
dev.off()
```

```
## pdf
##   2
```

The most damaging events for crops are drought (first position) and floods. Flood is also the most destroying event for properties couse damages of the highest value.