DATA ANALYSIS OF HEALTH AND ECONOMIC IMPACT OF METEOROLOGICAL EVENTS REGISTERED BY THE NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION'S DATABASE

Christian Jay

2022-07-06

SYNOPSIS

This data analysis aimed at defining what are the most destructive events regarding health and economic impact in the USA. Such impact will be measured with health variables (injuries and fatalities), and economic variables (property and crop damage). The database tracks adverse weather events during from 1950 to 2011. Tables compiling the top ten more devastating events and bar plots of the quantitative impact in USA population and economics will be displayed.

DATA PROCESSING

Before starting the analysis, we need to first load the packages required to process and plot the results obtained from the data set analysis. By using the function library(), we load the packages ggplot2, dplyr and reshape2.

We now proceed to read and explore the data set, located in the working directory set.

```
getwd()
```

```
## [1] "A:/Project Y06_Pre-Clinical Experiments/R Course/homework/Week 4 Course Project 2/Storm-Week4"
```

```
data <- read.csv(bzfile("repdata_data_StormData.csv.bz2"))
```

For exploring the data set we can use different functions such as head() or tails(), names(), colnames() and rownames(). But the simplest and more descriptive would be str(), which allow us to get a preliminary overview of the data set.

We see that the data set consists on 902.297 observations = rows and 37 variables = columns, with these containing both numerical and character data.

RESULTS

1) Across the United States, which types of events are most harmful with respect to population health?

The main variables regarding population health are injuries and fatalities. We have to aggregate both health variables to the evtype variable, corresponding to the event type (ej. winter snow). In order to do so we use the aggregate( ) function, we arrange( ) in descending order of the variable injuries (ej. max injuries to 0 injuries) using desc( ), and select only those events that have > 0 injuries records (rows 0:158 in total_injuries subset). To know, for example, the ten most devastating events regarding injuries we can just evaluate head(total_inj, 10). Same code for fatalities but only including rows 0:168 in total_fat, which are the rows corresponding to the events with > 0 fatalities.

1.1) INJURIES: Top 10

```
total_inj <- aggregate(INJURIES~EVTYPE, data, sum)
total_inj <- total_inj[total_inj$INJURIES > 0,]
total_inj <- arrange(total_inj, desc(INJURIES))
table_top10_inj <- head(total_inj, 10)
```

1.2) FATALITIES: Top 10

```
total_fat <- aggregate(FATALITIES~EVTYPE,data, sum)
total_fat <- total_fat[total_fat$FATALITIES > 0,]
total_fat <- arrange(total_fat, desc(FATALITIES))
table_top10_fat <- head(total_fat, 10)
```

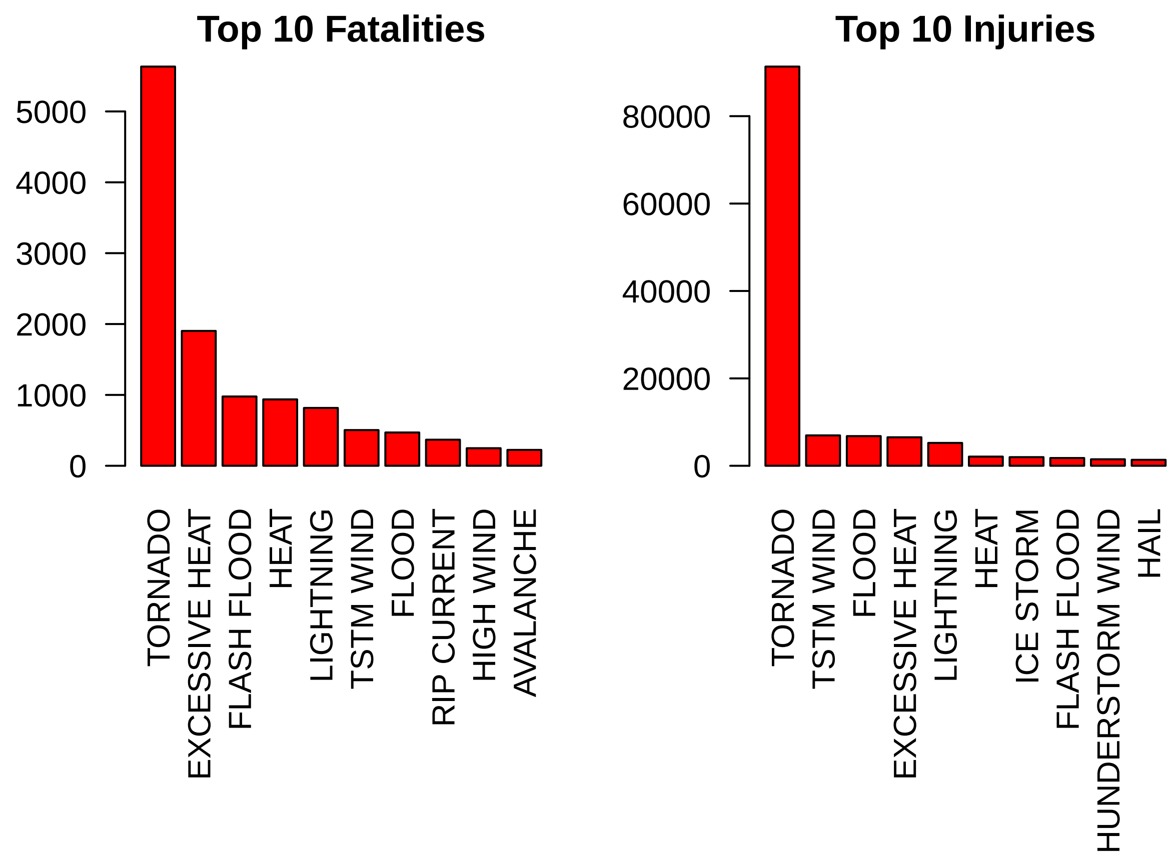
As we can see, tornadoes are the most destructive event regarding health impact. We can obtain the percentages regarding the sum of the injuries and fatalities of the top ten more destructive events knowing tornadoes are the top 1 in their data subset total_fat and total_inj.

```
torn_fat <- total_fat[1,2]
# Tornadoes account for 5633 fatalities
sum_fat_10 <- sum(head(total_fat["FATALITIES"], 10))
# The sum of the top 10 events is 12081
torn_perc_fat <- (torn_fat*100) / sum_fat_10
# Tornadoes account for more than 46% of fatalities of the top 10

torn_inj <- total_inj[1,2]
# Tornadoes account for 91346 injuries
sum_inj_10 <- sum(head(total_inj["INJURIES"], 10))
# The sum of the top 10 events is 125548
torn_perc_inj <- (torn_inj*100) / sum_inj_10
# Tornadoes account for more than 72% of injuries of the top 10
```

91246 of 140528 total injuries are caused by tornadoes, as well as 5633 of total 15145 fatalities. In other words, tornadoes account for 72.8% of injuries and 46.6% of fatalities in the top 10 most destructive events.

In order to better understand the contribution of tornadoes in this matter, we will plot these results into graphical representation. We perform two side-by-side bar plots of the top five injuries and fatalities events from the top ten ones. After optimizing the best margin and size configuration, we obtain the two-plot for injuries and fatalities together. This object will be PLOT 1, FIGURES 1 and 2.

```
par(mfrow=c(1,2), mar=c(10,4,2,1))
barplot(table_top10_fat$FATALITIES, names.arg=table_top10_fat$EVTYPE, las=2, col="red", ylab=, main="To
barplot(table_top10_inj$INJURIES, names.arg=table_top10_inj$EVTYPE, las=2, col="red", ylab= , main="Top
```



2) Across the United States, which types of events are most harmful with respect to economics?

Variables regarding economics are defined as property damage and crop damage. Same as with health impact, we use the aggregate( ), arrange( ) in descending order using desc( ) function, and select only those events that have > 0 damaging records (rows 0:404 in prop_dam subset and rows 0:136 in crop_dam subset). To know the ten most devastating events we use the function head(), coding head(prop_dam, 10) and head(crop_dam, 10) as before.

However, this time we need to transform some data due to the format in which it appears. As things get complicated in this section, we remove the variables we are not actually using, obtaining the same number of observations but addressing only 7 variables.

```
variables <- c("EVTYPE","FATALITIES","INJURIES","PROPDMG", "PROPDMGEXP","CROPDMG","CROPDMGEXP")
data <- data[variables]
str(data)
```

```
## 'data.frame':    902297 obs. of  7 variables:
##  $ EVTYPE    : chr  "TORNADO" "TORNADO" "TORNADO" "TORNADO" ...
##  $ FATALITIES: num  0 0 0 0 0 0 0 0 1 0 ...
##  $ INJURIES  : num  15 0 2 2 2 6 1 0 14 0 ...
##  $ PROPDMG   : num  25 2.5 25 2.5 2.5 2.5 2.5 2.5 25 25 ...
```

```
##  $ PROPDMGEXP: chr  "K" "K" "K" "K" ...
##  $ CROPDMG   : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ CROPDMGEXP: chr  "" "" "" "" ...
```

2.0) Exponential transforms to convert the exponent columns into numeric data

First we need to identify the different exp symbols and assign values for the property exponent to these symbols. For this process we just use the unique() function to the variables columns and assign the quantitative value.

```
unique(data$PROPDMGEXP)
```

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

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

data$PROPEXP[data$PROPDMGEXP == "K"] <- 1000
data$PROPEXP[data$PROPDMGEXP == "M"] <- 1e+06
data$PROPEXP[data$PROPDMGEXP == ""] <- 1
data$PROPEXP[data$PROPDMGEXP == "B"] <- 1e+09
data$PROPEXP[data$PROPDMGEXP == "m"] <- 1e+06
data$PROPEXP[data$PROPDMGEXP == "0"] <- 1
data$PROPEXP[data$PROPDMGEXP == "5"] <- 1e+05
data$PROPEXP[data$PROPDMGEXP == "6"] <- 1e+06
data$PROPEXP[data$PROPDMGEXP == "4"] <- 10000
data$PROPEXP[data$PROPDMGEXP == "2"] <- 100
data$PROPEXP[data$PROPDMGEXP == "3"] <- 1000
data$PROPEXP[data$PROPDMGEXP == "h"] <- 100
data$PROPEXP[data$PROPDMGEXP == "7"] <- 1e+07
data$PROPEXP[data$PROPDMGEXP == "H"] <- 100
data$PROPEXP[data$PROPDMGEXP == "1"] <- 10
data$PROPEXP[data$PROPDMGEXP == "8"] <- 1e+08
```

We also assign '0' to invalid exponent data and calculating the property damage value.

```
data$PROPEXP[data$PROPDMGEXP == "+"] <- 0
data$PROPEXP[data$PROPDMGEXP == "-"] <- 0
data$PROPEXP[data$PROPDMGEXP == "?"] <- 0

data$PROPDMG <- data$PROPDMG * data$PROPEXP
```

Now we do the same for the crop damage: detect the different exponential symbols and assigning values for the crop exp data.

```
unique(data$CROPDMGEXP)
```

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

```
#""  "M" "K" "m" "B" "?" "0" "k" "2"

data$CROPEXP[data$CROPDMGEXP == "M"] <- 1e+06
data$CROPEXP[data$CROPDMGEXP == "K"] <- 1000
data$CROPEXP[data$CROPDMGEXP == "m"] <- 1e+06
data$CROPEXP[data$CROPDMGEXP == "B"] <- 1e+09
data$CROPEXP[data$CROPDMGEXP == "0"] <- 1
data$CROPEXP[data$CROPDMGEXP == "k"] <- 1000
data$CROPEXP[data$CROPDMGEXP == "2"] <- 100
data$CROPEXP[data$CROPDMGEXP == ""] <- 1
```

Assigning '0' to invalid exponent data and calculating the crop damage.

```
data$CROPEXP[data$CROPDMGEXP == "?"] <- 0
data$CROPDMG <- data$CROPDMG * data$CROPEXP
```

After these data editions, our data set contain the same number of observations but now has 9 variables. We proceed to calculate property and crop damage as we did in the previous section, so we recycle the same code.

2.1) PROPERTY DAMAGE: Top 10 events

```
total_prop <- aggregate(PROPDMG~EVTYPE, data, sum)
total_prop <- total_prop[total_prop$PROPDMG > 0,]
total_prop <- arrange(total_prop, desc(PROPDMG))
table_top10_prop <- head(total_prop, 10)
```

2.2) CROP DAMAGE: Top 10 events

```
total_crop <- aggregate(CROPDMG~EVTYPE, data, sum)
total_crop <- total_crop[total_crop$CROPDMG > 0,]
total_crop <- arrange(total_crop, desc(CROPDMG))
table_top10_crop <- head(total_crop, 10)
```

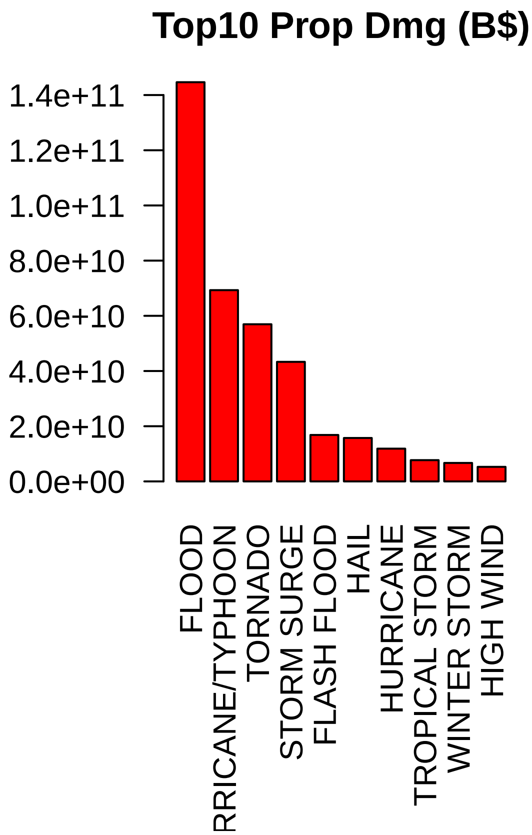
We conclude that floods are the most destructive event regarding prop damage, while droughts (and floods in second place) are the most destructive regarding crop damage. We can obtain the percentages regarding the sum of the prop and crop of the top ten more destructive events knowing they are the top 1 in each table

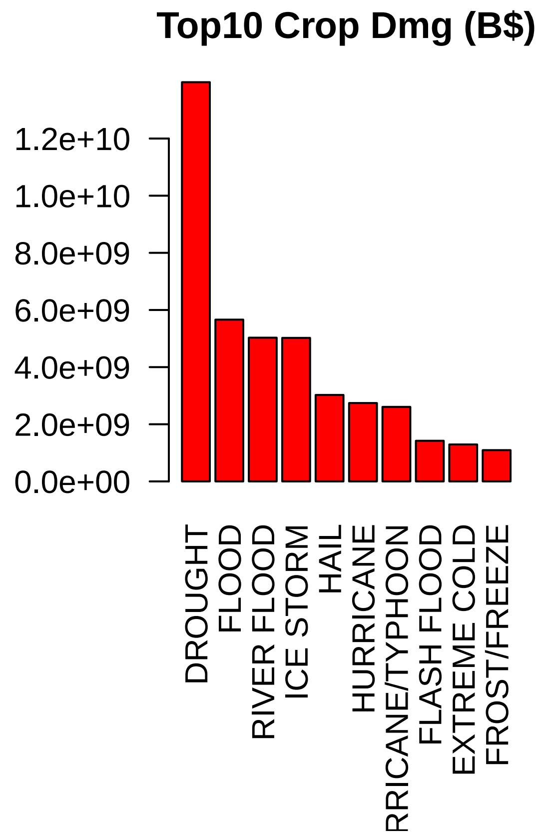
```
flood_prop <- total_prop[1,2]
# The property damage due to floods is 144.657.709.807$ (144.6B)
sum_prop_10 <- sum(head(total_prop["PROPDMG"], 10))
# The sum of property damage caused by the top ten most destructive events is 378.323.160.985$ (378B)
flood_perc_prop <- (total_prop[1,2]*100) / sum_prop_10
# Floods account for more than 38% of the sum of property damage in the top ten most destructive events

drou_crop <- total_crop[1,2]
# The crop damage due to droughts is 13.972.566.000$ (13.9B)
sum_crop_10 <- sum(head(total_crop["CROPDMG"], 10))
# The sum of crop damage caused by the top 10 mot destructive event is 41.870.220.323$ (41.8B)
drou_perc_crop <- (total_crop[1,2]*100) / sum_crop_10
# Droughts account for more than 33% of the sum of crop damage in the top ten most destructive events
```

144.657.709.807$ (144.6B) of 378.323.160.985$ (378B) total property damage are caused by floods, while 13.972.566.000$ (13.9B) of 41.870.220.323$ (41.8B) total crop damage are caused by droughts. In other words, floods account for 38.2% of prop damage and droughts 33.7% of crop damage.

In order to better understand the contribution of floods and droughts in this matter, we will plot these results into graphics as we did before with the health impact. This object will be PLOT 2, FIGURES 3 and 4.

```
par(mfrow=c(1,2), mar=c(9,5,3,2))
barplot(table_top10_prop$PROPDMG,names.arg=table_top10_prop$EVTYPE,las=2,col="red", ylab= , main="Top10
barplot(table_top10_crop$CROPDMG,names.arg=table_top10_crop$EVTYPE,las=2,col="red", ylab= , main="Top10
```





MAIN CONCLUSIONS

1) Regarding public health, tornadoes are the most devastating event for the most account of injuries and fatalities.

2) Regarding economic impact, floods account for the greatest property damage, while droughts account for the greatest crop damage, followed by foods in the second place.

Tornadoes, floods and droughts are concluded to be the most destructive events analyzing both the top ten most devastating events on health and economic impact.

Christian Jay (cjay@medlumics.com)