Analysis on the effects of storms and other severe weather events on public health and economy (61-year time span)

Summary:

Storms and other severe weather events can cause both public health and economic problems for communities and municipalities. Many severe events can result in fatalities, injuries, and property damage. The present analysis collects data from the U.S. National Oceanic and Atmospheric Administration's (NOAA) storm database for the years 1950 to 2011 and addresses two issues:

- Which types of events are most harmful to population health?
- Which types of events have the greatest economic consequences?

This preliminary analysis, based only in the extracted data without content modifications, yields the following conclusions:

1. The deadliest event was produced from a heat wave in Illinois on July 16th 1995, causing 583 deaths. Likewise, the most injuries were produced from a tornado in Texas, causing 1700 injuries (date undetermined).
2. Regarding the negative impact of weather events in general for human health, the biggest contributors are, by far, the temperature (heat, or excess thereof) and the wind (tornados/hurricanes), followed in a distant third/fourth place by floods and wildfires.
3. Regarding economic impact, the first tier was a flood in California on January 1st 2006, as an outlier with 115 billion, but most were strong-wind related events (8 out of the top 15), followed by storms.
4. Also couting towards economic negative impact but referring to crop damage, there was a tie for first event overall of a rivel flood and an ice storm (5 billion each), and the runner ups were more diverse, includind huricanes, droughts and extreme cold.

All the above conclusions are supported by the data here below.

```
##
## 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
```

```
##
## Attaching package: 'lubridate'
```

```
## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union
```

Data Processing:

Data was downloaded directly from the NOAA web site: Weather Data for Study (https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2).

The only modifications to original data were to reduce the variables to the interesting ones and changing the END_DATE variable to Date format for better processing. The first 10 rows are shown for descriptive purposes:

```
##    TIME_ZONE COUNTYNAME STATE  EVTYPE END_DATE FATALITIES INJURIES PROPDMG
## 1        CST     MOBILE    AL TORNADO     <NA>          0       15    25.0
## 2        CST    BALDWIN    AL TORNADO     <NA>          0        0     2.5
## 3        CST    FAYETTE    AL TORNADO     <NA>          0        2    25.0
## 4        CST    MADISON    AL TORNADO     <NA>          0        2     2.5
## 5        CST    CULLMAN    AL TORNADO     <NA>          0        2     2.5
## 6        CST LAUDERDALE    AL TORNADO     <NA>          0        6     2.5
## 7        CST     BLOUNT    AL TORNADO     <NA>          0        1     2.5
## 8        CST TALLAPOOSA    AL TORNADO     <NA>          0        0     2.5
## 9        CST TUSCALOOSA    AL TORNADO     <NA>          1       14    25.0
## 10       CST    FAYETTE    AL TORNADO     <NA>          0        0    25.0
##    PROPDMGEXP CROPDMG CROPDMGEXP
## 1           K       0
## 2           K       0
## 3           K       0
## 4           K       0
## 5           K       0
## 6           K       0
## 7           K       0
## 8           K       0
## 9           K       0
## 10          K       0
```

The programming language used was R, particularly packages “knitr”, “dplyr”,“lubridate” and “ggplot2”, to produce this RPubs publication out of original R markdown file. More specifics on the R version follow:

```
sessionInfo()
```

```
## R version 3.6.3 (2020-02-29)
## Platform: x86_64-w64-mingw32/x64 (64-bit)
## Running under: Windows 10 x64 (build 19041)
##
## Matrix products: default
##
## locale:
## [1] LC_COLLATE=English_United States.1252
## [2] LC_CTYPE=English_United States.1252
## [3] LC_MONETARY=English_United States.1252
## [4] LC_NUMERIC=C
## [5] LC_TIME=English_United States.1252
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
## [1] lubridate_1.7.9 knitr_1.29      ggplot2_3.3.2   dplyr_1.0.0
##
## loaded via a namespace (and not attached):
##  [1] Rcpp_1.0.5       magrittr_1.5     munsell_0.5.0    tidyselect_1.1.0
##  [5] colorspace_1.4-1 R6_2.4.1         rlang_0.4.6      stringr_1.4.0
##  [9] tools_3.6.3      grid_3.6.3       gtable_0.3.0     xfun_0.15
## [13] withr_2.2.0      htmltools_0.5.0  ellipsis_0.3.1   digest_0.6.25
## [17] tibble_3.0.2     lifecycle_0.2.0  crayon_1.3.4     purrr_0.3.4
## [21] vctrs_0.3.1      glue_1.4.1       evaluate_0.14    rmarkdown_2.3
## [25] stringi_1.4.6    compiler_3.6.3   pillar_1.4.6     generics_0.0.2
## [29] scales_1.1.1     pkgconfig_2.0.3
```

The downloaded data set showing the original http address is shown for more clarity:
download.file("https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2 (https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2)", "Dataset.csv") Dataset <- read.csv("Dataset.csv")

Data Analysis:

It is significant that, from the data, about 27% of the dates are missing, maybe never recorded, particularly for the initial years of the study, up to 1993:

```
sum(is.na(Dataset2$END_DATE))/nrow(Dataset2)*100
```

```
## [1] 26.97682
```

Analysis for effects on human health:

The most significant variables to be considered as affecting human health are the number of injuries and fatalities. In the period of study (1950 through 2011) a very high percentage of the records showed zero ocurrences:

```
sum(Dataset2$INJURIES==0)/nrow(Dataset2)*100
```

```
## [1] 98.04898
```

```
sum(Dataset2$FATALITIES==0)/nrow(Dataset2)*100
```

```
## [1] 99.22708
```

Which is, in essence, good news.

As a starting reference, the total numer of deaths and injured are the following:

```
Dataset2 %>% summarise(Total_Injured = sum(INJURIES),Total_Fatalities = sum(FATALITIES))
```

```
##   Total_Injured Total_Fatalities
## 1        140528            15145
```

Also as interesting data, I'll next identify the deadliest and the most harmful ocurrences, respectively:

```
Dataset2[Dataset2$FATALITIES==max(Dataset2$FATALITIES),3:7]
```

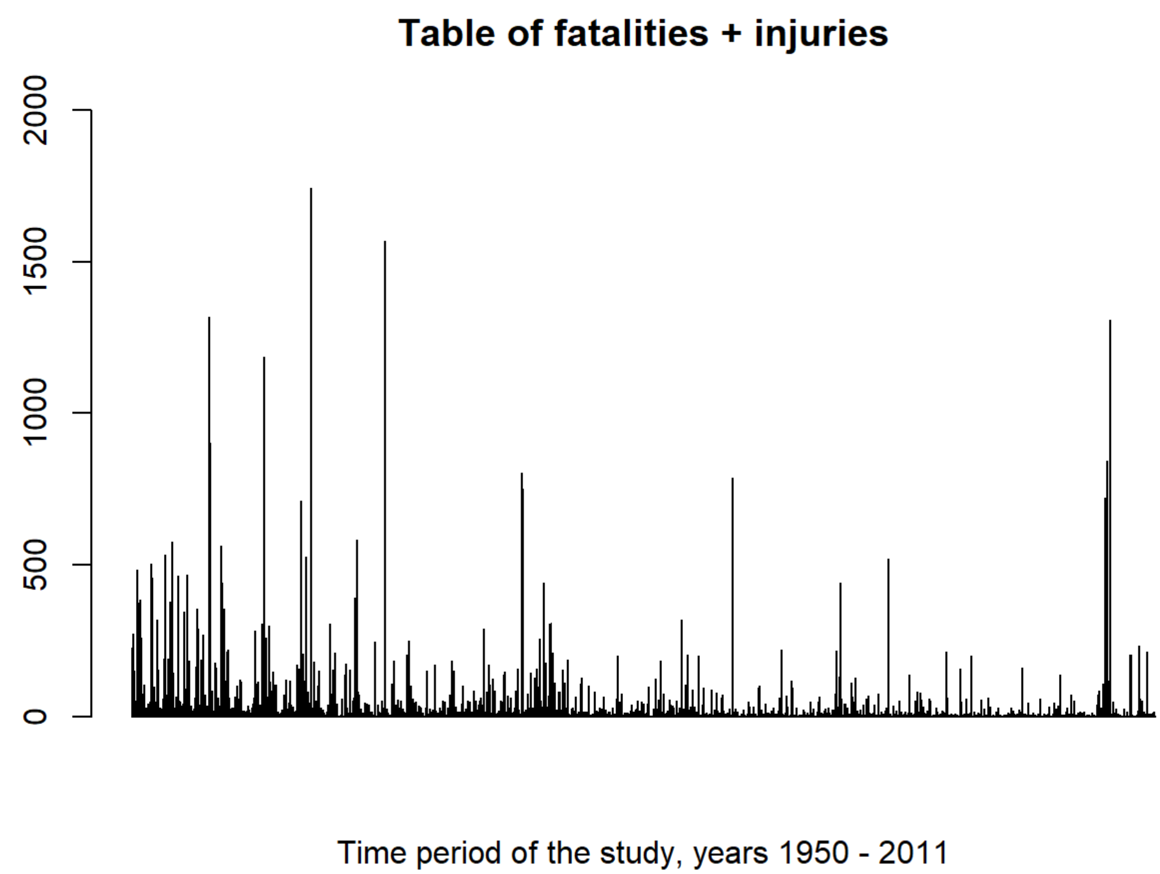
```
##        STATE EVTYPE   END_DATE FATALITIES INJURIES
## 198704    IL   HEAT 1995-07-16        583        0
```

```
Dataset2[Dataset2$INJURIES==max(Dataset2$INJURIES),3:7]
```

```
##        STATE  EVTYPE END_DATE FATALITIES INJURIES
## 157885    TX TORNADO     <NA>         42     1700
```

It follows from the previous data that the deadliest event was produced from a heat wave in Illinois on July 16th 1995, causing 583 deaths. Likewise, the most injuries were produced from a tornado in Texas, causing 1700 injuries (date undetermined).

To have a perspective on the development of all fatalities and injuries in the period of study from all causes, following please find a bar plot including all the weather factors:



From the grapgh above, the distribution of the events is irregular with noticiable spikes in the period, about 6 - 8 that are worth looking into specifically.

I will now sort the data based firstly on the number of fatalities and secondly on the number of injuries to determine the weather event types that mostly affect human health in this scenario:

```
Dataset2 <- Dataset2 %>% arrange(desc(FATALITIES,desc(INJURIES)))
Dataset2[1:50,c(4,6:7)]
```

```
##                        EVTYPE FATALITIES INJURIES
## 1                        HEAT        583        0
## 2                     TORNADO        158     1150
## 3                     TORNADO        116      785
## 4                     TORNADO        114      597
## 5              EXCESSIVE HEAT         99        0
## 6                     TORNADO         90     1228
## 7                     TORNADO         75      270
## 8              EXCESSIVE HEAT         74      135
## 9              EXCESSIVE HEAT         67        0
## 10                    TORNADO         57      504
## 11               EXTREME HEAT         57        0
## 12                    TORNADO         50      325
## 13             EXCESSIVE HEAT         49        0
## 14             EXCESSIVE HEAT         46       18
## 15                    TORNADO         44      800
## 16                    TORNADO         42     1700
## 17             EXCESSIVE HEAT         42      397
## 18             EXCESSIVE HEAT         42        0
## 19                    TORNADO         38      270
## 20                    TORNADO         37      176
## 21                    TORNADO         36     1150
## 22                    TORNADO         34      350
## 23                    TORNADO         33      500
## 24                  HEAT WAVE         33        0
## 25             EXCESSIVE HEAT         33        0
## 26                    TORNADO         32      258
## 27             EXCESSIVE HEAT         32        0
## 28                    TSUNAMI         32      129
## 29                    TORNADO         31      252
## 30                    TORNADO         31      252
## 31                    TORNADO         31      257
## 32                    TORNADO         30      411
## 33                    TORNADO         30      121
## 34             EXCESSIVE HEAT         30        0
## 35                    TORNADO         29      180
## 36                    TORNADO         29      350
## 37  UNSEASONABLY WARM AND DRY         29        0
## 38                       HEAT         27        0
## 39                    TORNADO         27       12
## 40                    TORNADO         27        0
## 41                    TORNADO         26      500
## 42                    TORNADO         25      200
## 43                    TORNADO         25      342
## 44 TORNADOES, TSTM WIND, HAIL         25        0
## 45                  HEAT WAVE         25        0
## 46                    TORNADO         25      145
## 47                    TORNADO         24      410
## 48             EXCESSIVE HEAT         24       60
## 49             EXCESSIVE HEAT         24        0
## 50             EXCESSIVE HEAT         24       40
```

From the data, it becomes apparent that the biggest contributors are, by far, the temperature (heat, or excess thereof) and the wind (tornados/hurricanes), followed in a distant third/fourth place by floods and wildfires.

If we calculate the proportion of health damage produced only by heat:

```
Dataset2 %>% filter(grepl("HEAT",EVTYPE, ignore.case = T)) %>% summarise(Total_Injured_from_heat
= sum(INJURIES),Total_Fatalities_from_heat = sum(FATALITIES))
```

```
##   Total_Injured_from_heat Total_Fatalities_from_heat
## 1                    9224                       3138
```

Therefore about 15% of the all injuries were caused by some kind of heat factor and around 26% of all fatalities as well, which is quite significant.

If we do the same analysis for our second contributor(tornados), we obtain:

```
Dataset2 %>% filter(grepl("TORNADO|HURRICANE|WIND",EVTYPE, ignore.case = T)) %>% summarise(Total
_Injured_from_tornadoes = sum(INJURIES),Total_Fatalities_from_tornadoes = sum(FATALITIES))
```

```
##   Total_Injured_from_tornadoes Total_Fatalities_from_tornadoes
## 1                       104233                            7220
```

Therefore a whooping 74% of the all injuries were caused by some kind of tornado/hurricane and around 48% of all fatalities, **which sets this wheather factor apart as the most harmful for the human being in the period of study.**

Continuing on down the table we find the next tiers (floods,storms,cold). the following chart shows the top 5 and its totals:

Analysis for effects on economy:

For this anylisis it makes sense to use the “damage” related data in the provided data set, namely:

- Property Damage (PROPDMG).
- Property Damage exponent (PROPDMGEXP).
- Crop Damage Data (CROPDMG).
- Crop Damage Data exponent (CROPDMGEXP).

There is a thorough study conducted a few years ago regarding the exponent value for the Damage data : exponent values explained (https://github.com/flyingdisc/RepData_PeerAssessment2/blob/master/how-to-handle-PROPDMGEXP.md), which attempts to explain the possible values of the variables in PROPDMGEXP and CROPDMGEXP that affect the values in PROPDMG and CROPDMG respectively:

```
levels(Dataset$PROPDMGEXP)
```

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

```
levels(Dataset$CROPDMGEXP)
```

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

In brief:
H,h,K,k,M,m,B,b,+,-,?,0,1,2,3,4,5,6,7,8, and blank-character

H,h = hundreds = 100

K,k = kilos = thousands = 1,000

M,m = millions = 1,000,000

B,b = billions = 1,000,000,000

(+) = 1

(-) = 0

(?) = 0

black/empty character = 0

numeric 0..8 = 10

A quick calculation yields that real values for PROPDMG and CROPDMG only show in about 5% and 11% of all data respectively, therefore special attention must be paid to values in the millions or billions.

On the other hand, contribution of the numerical and other values besides the exponentials (B,M,K) values are negligible and for the purpose of the present study they could be omitted:

```
## [1] "0.0348%"
```

```
## [1] "0.003%"
```

The total damage produced for these above also turned out minimal. In any case, I will not disregard almost any contributor, to make the conclusions more reliable.

I will proceed to subset the property damage (PROPDM) and crop damage (CROPDMG) to determine the impact on economy, first setting the total dollar value of each variable according the exponencial value determined for PROPDMGEXP and CROPDMGEXP, as explained above, correcting for billions (B), millions (M) and thousands (K), expecting to see the greatest contributions from those observations.

Let’s see the observations with 9-figure amounts, I created a new column as “Total_PROP_mill”, accounting for total property damage in millions of US dollars, arranged in descending order:

```
Data_B <- Dataset2 %>% filter(PROPDMGEXP=="B") %>% mutate(Total_PROP_mill = PROPDMG*1000) %>% ar
range(desc(Total_PROP_mill))
Data_B[,c(3:5,12)]
```

```
##    STATE                     EVTYPE   END_DATE Total_PROP_mill
## 1     CA                      FLOOD 2006-01-01          115000
## 2     LA                STORM SURGE 2005-08-29           31300
## 3     LA          HURRICANE/TYPHOON 2005-08-29           16930
## 4     MS                STORM SURGE 2005-08-29           11260
## 5     FL          HURRICANE/TYPHOON 2005-10-24           10000
## 6     MS          HURRICANE/TYPHOON 2005-08-29            7350
## 7     MS          HURRICANE/TYPHOON 2005-08-29            5880
## 8     FL          HURRICANE/TYPHOON 2004-08-13            5420
## 9     TX             TROPICAL STORM 2001-06-10            5150
## 10    AL               WINTER STORM 1993-03-13            5000
## 11    IL                RIVER FLOOD       <NA>            5000
## 12    FL          HURRICANE/TYPHOON 2004-09-05            4830
## 13    TX           STORM SURGE/TIDE 2008-09-14            4000
## 14    FL          HURRICANE/TYPHOON 2004-09-16            4000
## 15    LA          HURRICANE/TYPHOON 2005-09-24            4000
## 16    ND                      FLOOD 1997-04-23            3000
## 17    NC                  HURRICANE 1999-09-16            3000
## 18    MO                    TORNADO 2011-05-22            2800
## 19    LA  HEAVY RAIN/SEVERE WEATHER 1995-05-10            2500
## 20    AL          HURRICANE/TYPHOON 2004-09-16            2500
## 21    FL             HURRICANE OPAL 1995-10-04            2100
## 22    TX          HURRICANE/TYPHOON 2005-09-24            2090
## 23    TN                      FLOOD 2011-05-31            2000
## 24    AZ                       HAIL 2010-10-05            1800
## 25    PR                  HURRICANE 1998-09-22            1700
## 26    FL TORNADOES, TSTM WIND, HAIL 1993-03-13            1600
## 27    AL                    TORNADO 2011-04-27            1500
## 28    TN                      FLOOD 2010-05-04            1500
## 29    NM           WILD/FOREST FIRE 2000-05-31            1500
## 30    FL          HURRICANE/TYPHOON 2005-07-10            1500
## 31    FL                  HIGH WIND 2004-08-13            1300
## 32    TX        SEVERE THUNDERSTORM       <NA>            1200
## 33    CA                   WILDFIRE 2003-10-31            1040
## 34    AL                    TORNADO 2011-04-27            1000
## 35    FL             HURRICANE OPAL       <NA>            1000
## 36    AL                FLASH FLOOD 2003-05-08            1000
## 37    AL          HURRICANE/TYPHOON 2005-08-29            1000
## 38    TX                  HURRICANE 2008-09-14            1000
## 39    MS                      FLOOD 2011-05-31            1000
## 40    AL  HURRICANE OPAL/HIGH WINDS 1995-10-05             100
```

The first one in the list corresponds to a flood in California on January 1st 2006 (115 billion). Then three significant storm related events follow in August of the same year, presumably originated by Hurricane Katrina (confirmed in the same data set).

Now the same exercise for the crop damage, named “Total_CROP_mill”, 9-figure events:

```
Data_BC <- Dataset2 %>% filter(CROPDMGEXP=="B") %>% mutate(Total_CROP_mill = CROPDMG*1000) %>% a
rrange(desc(Total_CROP_mill))
Data_BC[,c(3:5,12)]
```

```
##   STATE            EVTYPE   END_DATE Total_CROP_mill
## 1    IL       RIVER FLOOD       <NA>            5000
## 2    MS         ICE STORM 1994-02-10            5000
## 3    MS HURRICANE/TYPHOON 2005-08-29            1510
## 4    TX           DROUGHT 2006-01-31            1000
## 5    IA           DROUGHT 1995-08-31             500
## 6    AL              HEAT       <NA>             400
## 7    IA            FREEZE 1995-09-22             200
## 8    GA           DROUGHT 2011-09-30               0
## 9    GA           DROUGHT 2011-09-30               0
```

Herefore, a river flood pops up as number one in Illinois, date undetermined, in a tie with an ice storm in the state of Mississippi on the shown date (5 billion each).

The next group in order of magnitud are the millions, I will then subset also those observations for property and crops, again the total will show in millions USD (this time only first 20 rows shown due to the quantity):

```
Data_M <- Dataset2 %>% filter(PROPDMGEXP=="M") %>% mutate(Total_PROP_mill = PROPDMG) %>% arrange
(desc(Total_PROP_mill))
head(Data_M[,c(3:5,12)],20)
```

```
##    STATE            EVTYPE   END_DATE Total_PROP_mill
## 1     FL         HIGH WIND 2004-08-13          929.00
## 2     AZ              HAIL 2010-10-05          900.00
## 3     NC         HURRICANE 1996-09-06          792.15
## 4     IA             FLOOD 2008-06-30          750.00
## 5     OH      WINTER STORM 2008-03-08          750.00
## 6     OK THUNDERSTORM WIND 2008-06-05          750.00
## 7     FL         HIGH WIND 2004-09-26          702.00
## 8     AL           TORNADO 2011-04-27          700.00
## 9     LA THUNDERSTORM WIND 2009-05-10          700.00
## 10    CA          WILDFIRE 2003-10-31          696.40
## 11    IA           DROUGHT 2003-08-31          645.15
## 12    NV             FLOOD 1997-01-17          640.00
## 13    FL HURRICANE/TYPHOON 2004-09-05          621.00
## 14    CA        WILD FIRES       <NA>          619.00
## 15    MS         HURRICANE 1998-09-28          602.00
## 16    MN             FLOOD 1997-04-23          600.00
## 17    CA  WILD/FOREST FIRE 2000-09-30          547.00
## 18    MD    TROPICAL STORM 2003-09-19          530.47
## 19    KY              HAIL 1998-04-16          510.00
## 20    VA HURRICANE/TYPHOON 2003-09-19          506.00
```

- For crops:

```
Data_MC <- Dataset2 %>% filter(CROPDMGEXP=="M") %>% mutate(Total_CROP_mill = CROPDMG) %>% arrang
e(desc(Total_CROP_mill))
head(Data_MC[,c(3:5,12)],20)
```

```
##    STATE           EVTYPE   END_DATE Total_CROP_mill
## 1     CA      EXTREME COLD 1998-12-27          596.00
## 2     IA           DROUGHT 2001-08-23          578.85
## 3     TX           DROUGHT 2000-11-30          515.00
## 4     OK           DROUGHT 1998-07-31          500.00
## 5     NC         HURRICANE 1999-09-16          500.00
## 6     PA           DROUGHT 1999-07-31          500.00
## 7     FL             FLOOD 2000-10-04          500.00
## 8     MO             FLOOD 2007-07-03          500.00
## 9     CA    EXCESSIVE HEAT 2006-07-27          492.40
## 10    NE           DROUGHT 2002-12-31          480.00
## 11    TX           DROUGHT 1998-12-31          450.00
## 12    FL HURRICANE/TYPHOON 2005-08-26          423.00
## 13    TX           DROUGHT 2001-12-31          420.00
## 14    NC         HURRICANE 1999-09-16          413.60
## 15    OK           DROUGHT 2000-08-31          399.84
## 16    GA           DROUGHT 2007-09-30          344.00
## 17    FL         HURRICANE 1999-10-15          338.00
## 18    IA           DROUGHT 2003-08-31          312.48
## 19    GA           DROUGHT 2000-06-30          306.72
## 20    PR         HURRICANE 1998-09-22          301.00
```

In this case extreme cold and drought took first and second places, in this order of magnitud(6 figures).

Following this initial procedure, I will subset the most significant events, to produce a graph representing the most harmul events for economy, in terms of lost property and crop loss, after normalizing the dollar figures (to millions) and creating one final data base with them. I will start from the original data set for more clarity and reproducibility:

```
DF0 <- read.csv("Dataset.csv",stringsAsFactors = F)
DF1 <- DF0[,c(8,25:28)]

i <- numeric()

for (i in 1:nrow(DF1)){
 if(DF1$PROPDMGEXP[i]=="B"){DF1$PROPDMG[i]=DF1$PROPDMG[i]*1000}
   if(DF1$PROPDMGEXP[i]=="M"){DF1$PROPDMG[i]=DF1$PROPDMG[i]}
     if(DF1$PROPDMGEXP[i]=="K"){DF1$PROPDMG[i]=DF1$PROPDMG[i]/1000}
       if(DF1$CROPDMGEXP[i]=="B"){DF1$CROPDMG[i]=DF1$CROPDMG[i]*1000}
         if(DF1$CROPDMGEXP[i]=="M"){DF1$CROPDMG[i]=DF1$CROPDMG[i]}
           if(DF1$CROPDMGEXP[i]=="K"){DF1$CROPDMG[i]=DF1$CROPDMG[i]/1000}
                     }
```

Looking at the first events, after ordering first by property damage and second by crop damage:

```
head(arrange(DF1,desc(PROPDMG),desc(CROPDMG)),20)[,c(1:2,4)]
```

```
##                       EVTYPE PROPDMG CROPDMG
## 1                      FLOOD  115000    32.5
## 2                STORM SURGE   31300     0.0
## 3          HURRICANE/TYPHOON   16930     0.0
## 4                STORM SURGE   11260     0.0
## 5          HURRICANE/TYPHOON   10000     0.0
## 6          HURRICANE/TYPHOON    7350     0.0
## 7          HURRICANE/TYPHOON    5880  1510.0
## 8          HURRICANE/TYPHOON    5420   285.0
## 9             TROPICAL STORM    5150     0.0
## 10               RIVER FLOOD    5000  5000.0
## 11              WINTER STORM    5000     0.0
## 12         HURRICANE/TYPHOON    4830    93.2
## 13         HURRICANE/TYPHOON    4000    25.0
## 14         HURRICANE/TYPHOON    4000     0.0
## 15          STORM SURGE/TIDE    4000     0.0
## 16                 HURRICANE    3000   500.0
## 17                     FLOOD    3000     0.0
## 18                   TORNADO    2800     0.0
## 19         HURRICANE/TYPHOON    2500    25.0
## 20 HEAVY RAIN/SEVERE WEATHER    2500     0.0
```

Now the other way around, ordering first by crop damage:

```
head(arrange(DF1,desc(CROPDMG),desc(PROPDMG)),20)[,c(1:2,4)]
```

```
##               EVTYPE  PROPDMG CROPDMG
## 1        RIVER FLOOD 5000.000 5000.00
## 2          ICE STORM    0.500 5000.00
## 3  HURRICANE/TYPHOON 5880.000 1510.00
## 4            DROUGHT    0.000 1000.00
## 5       EXTREME COLD    0.000  596.00
## 6            DROUGHT    0.000  578.85
## 7            DROUGHT    0.000  515.00
## 8          HURRICANE 3000.000  500.00
## 9              FLOOD  450.000  500.00
## 10             FLOOD    0.005  500.00
## 11           DROUGHT    0.000  500.00
## 12           DROUGHT    0.000  500.00
## 13           DROUGHT    0.000  500.00
## 14    EXCESSIVE HEAT    0.170  492.40
## 15           DROUGHT    0.000  480.00
## 16           DROUGHT    0.000  450.00
## 17 HURRICANE/TYPHOON  100.000  423.00
## 18           DROUGHT    0.000  420.00
## 19         HURRICANE  410.620  413.60
## 20              HEAT    0.000  400.00
```

Therefore it is safe to conclude that, for property damage, we confirm the flood as first, as an outlier with 115 billion, followed by strong-wind related events (8 out of the top 15), followed by storms.

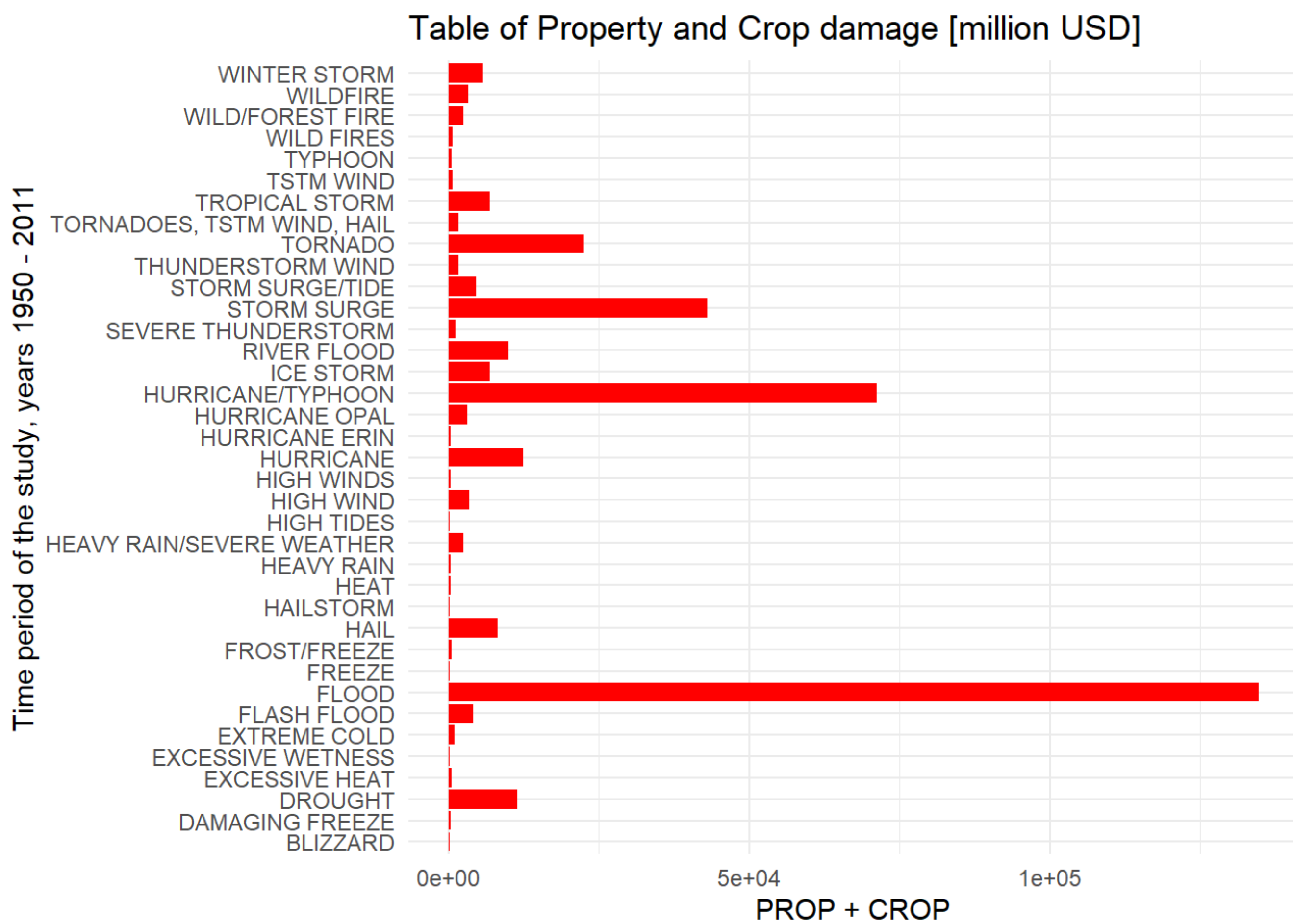
Final plot for damages assesment:

The data for the plots were filtered to show only around 40 top factors, for illustrative purposes:

```
Plot2 <- DF1 %>% filter(PROPDMG>100|CROPDMG>100) %>%
 group_by(EVTYPE) %>% summarise(PROP=sum(PROPDMG), CROP=sum(CROPDMG))
```

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

```
ggplot(Plot2, aes(x=EVTYPE,y=PROP+CROP))+
        geom_bar(stat = "identity", fill="red")+
     ggtitle("Table of Property and Crop damage [million USD]")+
     xlab("Time period of the study, years 1950 - 2011")+
     coord_flip()+
     theme_minimal()
```



Lastly, for comparative purposes, a graph showing only the damage to crops follows, keeping the same factors and monetary scale (millions USD):

```
ggplot(Plot2, aes(x=EVTYPE,y=CROP))+
 geom_bar(stat = "identity", fill="blue")+
 ggtitle("Table of Crop damage [million USD]")+
 xlab("Time period of the study, years 1950 - 2011")+
 coord_flip()+
 theme_minimal()
```

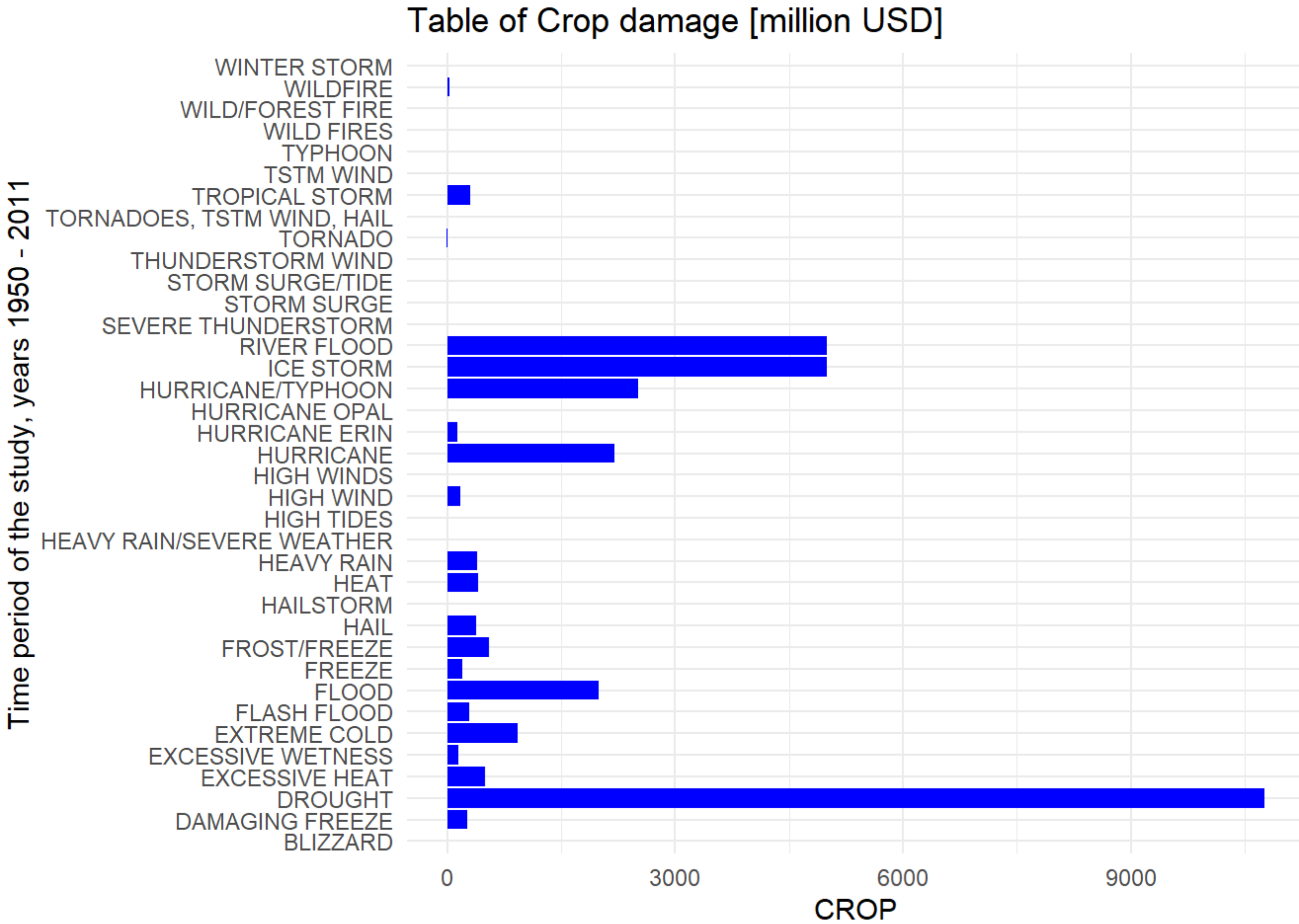


Table of Crop damage [million USD]
WINTER STORM
WILDFIRE
WILD/FOREST FIRE
WILD FIRES
TYPHOON
TSTM WIND
TROPICAL STORM
TORNADOES, TSTM WIND, HAIL
TORNADO
THUNDERSTORM WIND
STORM SURGE/TIDE
STORM SURGE
SEVERE THUNDERSTORM
RIVER FLOOD
ICE STORM
HURRICANE/TYPHOON
HURRICANE OPAL
HURRICANE ERIN
HURRICANE
HIGH WINDS
HIGH WIND
HIGH TIDES
HEAVY RAIN/SEVERE WEATHER
HEAVY RAIN
HEAT
HAILSTORM
HAIL
FROST/FREEZE
FREEZE
FLOOD
FLASH FLOOD
EXTREME COLD
EXCESSIVE WETNESS
EXCESSIVE HEAT
DROUGHT
DAMAGING FREEZE
BLIZZARD
Time period of the study, years 1950 - 2011
0
3000
6000
9000
CROP