Severe Weather Events and their effects

preethi

10/26/2019

Synopsis

In this report we look at data from the NOAA storm database and answer two questions

1) Across the United States, which types of events (as indicated by the EVTYPE variable) are most harmful with respect to population health?
2) Across the United States, which types of events have the greatest economic consequences?

##Brief description of the database

1) The events in the database start in the year 1950 and end in November 2019. In the earlier years of the database there are generally fewer events recorded, most likely due to a lack of good records. More recent years should be considered more complete.

2) The data is obtained from the National Weather Service and is publishehd by the National Oceanic and Atmospheric Administration (NOAA) in an attempt to record the occurence of weeather phenomena of sufficient intensity to cause loss of life, injuries, significant property damage, and/or disruption to commerce

Loading and processing the raw data

The data for this report is a part of the [Storm Events Database] (https://www.ncdc.noaa.gov/stormevents/)

The file for our purposes is a comma-seperated-value file compressed with bzip2. We will be using the data gathered from 1950 to Nov 2011 for our analysis.

```r
library(dplyr)
```

```
##
## 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
```

```r
library(ggplot2)
```

```
## Registered S3 methods overwritten by 'ggplot2':
##   method         from
##   [.quosures     rlang
##   c.quosures     rlang
##   print.quosures rlang
```

```r
if(!file.exists("StormData.csv.bz2")){
  download.file("https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2",
                "StormData.csv.bz2")
}

Stormdata <- read.csv("StormData.csv.bz2",header = TRUE)
```

We can now look at the dataset

```r
dim(Stormdata)
```

```
## [1] 902297     37
```

```r
head(Stormdata)
```

```
##   STATE__           BGN_DATE BGN_TIME TIME_ZONE COUNTY COUNTYNAME STATE
## 1       1  4/18/1950 0:00:00     0130       CST     97     MOBILE    AL
## 2       1  4/18/1950 0:00:00     0145       CST      3    BALDWIN    AL
## 3       1  2/20/1951 0:00:00     1600       CST     57    FAYETTE    AL
## 4       1   6/8/1951 0:00:00     0900       CST     89    MADISON    AL
## 5       1 11/15/1951 0:00:00     1500       CST     43    CULLMAN    AL
## 6       1 11/15/1951 0:00:00     2000       CST     77 LAUDERDALE    AL
##    EVTYPE BGN_RANGE BGN_AZI BGN_LOCATI END_DATE END_TIME COUNTY_END
## 1 TORNADO         0                                               0
## 2 TORNADO         0                                               0
## 3 TORNADO         0                                               0
## 4 TORNADO         0                                               0
## 5 TORNADO         0                                               0
## 6 TORNADO         0                                               0
##   COUNTYENDN END_RANGE END_AZI END_LOCATI LENGTH WIDTH F MAG FATALITIES
## 1         NA         0                      14.0   100 3   0          0
## 2         NA         0                       2.0   150 2   0          0
## 3         NA         0                       0.1   123 2   0          0
## 4         NA         0                       0.0   100 2   0          0
## 5         NA         0                       0.0   150 2   0          0
## 6         NA         0                       1.5   177 2   0          0
##   INJURIES PROPDMG PROPDMGEXP CROPDMG CROPDMGEXP WFO STATEOFFIC ZONENAMES
## 1       15    25.0          K       0
## 2        0     2.5          K       0
## 3        2    25.0          K       0
## 4        2     2.5          K       0
## 5        2     2.5          K       0
## 6        6     2.5          K       0
##   LATITUDE LONGITUDE LATITUDE_E LONGITUDE_ REMARKS REFNUM
## 1     3040      8812       3051       8806              1
## 2     3042      8755          0          0              2
## 3     3340      8742          0          0              3
## 4     3458      8626          0          0              4
## 5     3412      8642          0          0              5
## 6     3450      8748          0          0              6
```

```r
str(Stormdata)
```

```
## 'data.frame':    902297 obs. of  37 variables:
##  $ STATE__   : num  1 1 1 1 1 1 1 1 1 1 ...
##  $ BGN_DATE  : Factor w/ 16335 levels "1/1/1966 0:00:00",..: 6523 6523 4242 11116 2224 2224 2260 383
##  $ BGN_TIME  : Factor w/ 3608 levels "00:00:00 AM",..: 272 287 2705 1683 2584 3186 242 1683 3186 318
##  $ TIME_ZONE : Factor w/ 22 levels "ADT","AKS","AST",..: 7 7 7 7 7 7 7 7 7 7 ...
##  $ COUNTY    : num  97 3 57 89 43 77 9 123 125 57 ...
##  $ COUNTYNAME: Factor w/ 29601 levels "","5NM E OF MACKINAC BRIDGE TO PRESQUE ISLE LT MI",..: 13513 
##  $ STATE     : Factor w/ 72 levels "AK","AL","AM",..: 2 2 2 2 2 2 2 2 2 2 ...
##  $ EVTYPE    : Factor w/ 985 levels "   HIGH SURF ADVISORY",..: 834 834 834 834 834 834 834 834 834 8
##  $ BGN_RANGE : num  0 0 0 0 0 0 0 0 0 0 ...
```

```
##  $ BGN_AZI   : Factor w/ 35 levels "","  N"," NW",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ BGN_LOCATI: Factor w/ 54429 levels "","- 1 N Albion",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ END_DATE  : Factor w/ 6663 levels "","1/1/1993 0:00:00",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ END_TIME  : Factor w/ 3647 levels ""," 0900CST",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ COUNTY_END: num  0 0 0 0 0 0 0 0 0 0 ...
##  $ COUNTYENDN: logi  NA NA NA NA NA NA ...
##  $ END_RANGE : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ END_AZI   : Factor w/ 24 levels "","E","ENE","ESE",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ END_LOCATI: Factor w/ 34506 levels "","- .5 NNW",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ LENGTH    : num  14 2 0.1 0 0 1.5 1.5 0 3.3 2.3 ...
##  $ WIDTH     : num  100 150 123 100 150 177 33 33 100 100 ...
##  $ F         : int  3 2 2 2 2 2 2 1 3 3 ...
##  $ MAG       : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ 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: Factor w/ 19 levels "","-","?","+",..: 17 17 17 17 17 17 17 17 17 17 ...
##  $ CROPDMG   : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ CROPDMGEXP: Factor w/ 9 levels "","?","0","2",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ WFO       : Factor w/ 542 levels ""," CI","$AC",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ STATEOFFIC: Factor w/ 250 levels "","ALABAMA, Central",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ ZONENAMES : Factor w/ 25112 levels "","
##  $ LATITUDE  : num  3040 3042 3340 3458 3412 ...
##  $ LONGITUDE : num  8812 8755 8742 8626 8642 ...
##  $ LATITUDE_E: num  3051 0 0 0 0 ...
##  $ LONGITUDE_: num  8806 0 0 0 0 ...
##  $ REMARKS   : Factor w/ 436781 levels "","-2 at Deer Park\n",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ REFNUM    : num  1 2 3 4 5 6 7 8 9 10 ...
```

We can now subset the table with the parameters that we would like to look at. EVTYPE,F,MAG,FATALITIES,INJURIES,PRO and CROPDMGEXP

```
Stormdata_subset <- Stormdata[,c("EVTYPE","F","MAG","FATALITIES","INJURIES",
                                 "PROPDMG","PROPDMGEXP","CROPDMG", "CROPDMGEXP")]
```

As mentioned in [National Weather Service Storm Data Documentation] (https://d396qusza40orc.cloudfront.net/repdata%2Fpeer2_doc%2Fpd01016005curr.pdf) the value K corresponds to thousand, M for a million and B for a billion and these valid qualifiers are found in PROPDMGEXP and CROPDMGEXP. To make it less confusing and for initial analysis, we have created a new column total damage with the crop and property damage added and listed together. This was done by creating a function total damage which takes as its arguments PROPDMG,PROPDMGEXP, CROPDMG and CROPDMGEXP.

```
total_damage <- function(prop_dmg,prop_exp,crop_dmg,crop_exp){
 propdam <- 0
 cropdam <- 0

 if (prop_dmg > 0){
   if (tolower(prop_exp) == "h")
       propdam <- prop_dmg * 0
   if (tolower(prop_exp) == "k")
       propdam <- prop_dmg * 1000
   if  (tolower(prop_exp) == "m")
       propdam <- prop_dmg * 1000000
   if (tolower(prop_exp) == "b")
       propdam <- prop_dmg * 1000000000
  }
```

```
  if (crop_dmg > 0){
   if (tolower(crop_exp) == "h")
       propdam <- crop_dmg * 0
   if (tolower(crop_exp) == "k")
       propdam <- crop_dmg * 1000
   if  (tolower(crop_exp) == "m")
       propdam <- crop_dmg * 1000000
   if (tolower(crop_exp) == "b")
       propdam <- crop_dmg * 1000000000
  }
  return(propdam+cropdam)
}
Stormdata_subset$Totaldam <- mapply(total_damage,Stormdata_subset$PROPDMG,Stormdata_subset$PROPDMGEXP,
                                    Stormdata_subset$CROPDMG,Stormdata_subset$CROPDMGEXP)
```

Results

1) Across the United States, which types of events (as indicated by the EVTYPE variable) are most harmful with respect to population health?

To do this, we first subset the data based on weather pattern and then calculate the sum of injuries/fatalities for the respective patterns. This is the put forth in the form of a table and figure for the 10 most harmful weather patterns (since there is too many to show them all).

```
injuries <- Stormdata_subset %>% group_by(EVTYPE)%>%
  summarise(injuries = sum(INJURIES,na.rm = TRUE)) %>% as.data.frame() %>% rename(Weather_pattern=EVTYPE
top_ten_i <- head(subset(injuries[order(injuries$injury_count,decreasing = TRUE),]),10)
top_ten_i
```

```
##       Weather_pattern injury_count
## 834           TORNADO        91346
## 856         TSTM WIND         6957
## 170             FLOOD         6789
## 130    EXCESSIVE HEAT         6525
## 464         LIGHTNING         5230
## 275              HEAT         2100
## 427         ICE STORM         1975
## 153       FLASH FLOOD         1777
## 760 THUNDERSTORM WIND         1488
## 244              HAIL         1361
```

```
fatalities <- Stormdata_subset %>% group_by(EVTYPE)%>%
  summarise(fatalities = sum(FATALITIES,na.rm = TRUE)) %>% as.data.frame() %>% rename(Weather_pattern=E
top_ten_f <- head(subset(fatalities[order(fatalities$fatality_count,decreasing=TRUE),]),10)
top_ten_f
```

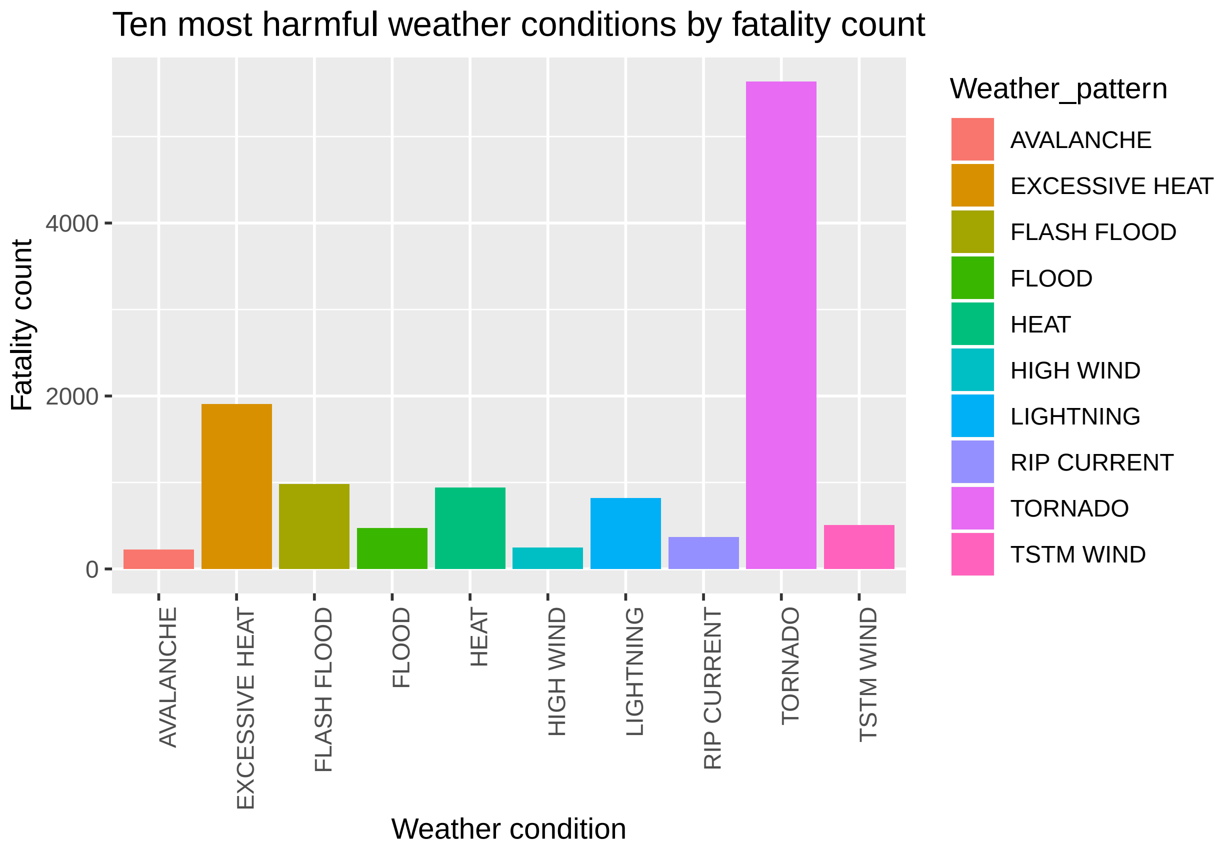
```
##     Weather_pattern fatality_count
## 834         TORNADO           5633
## 130  EXCESSIVE HEAT           1903
## 153     FLASH FLOOD            978
## 275            HEAT            937
## 464       LIGHTNING            816
## 856       TSTM WIND            504
## 170           FLOOD            470
## 585     RIP CURRENT            368
```

```
## 359        HIGH WIND              248
## 19         AVALANCHE              224
```

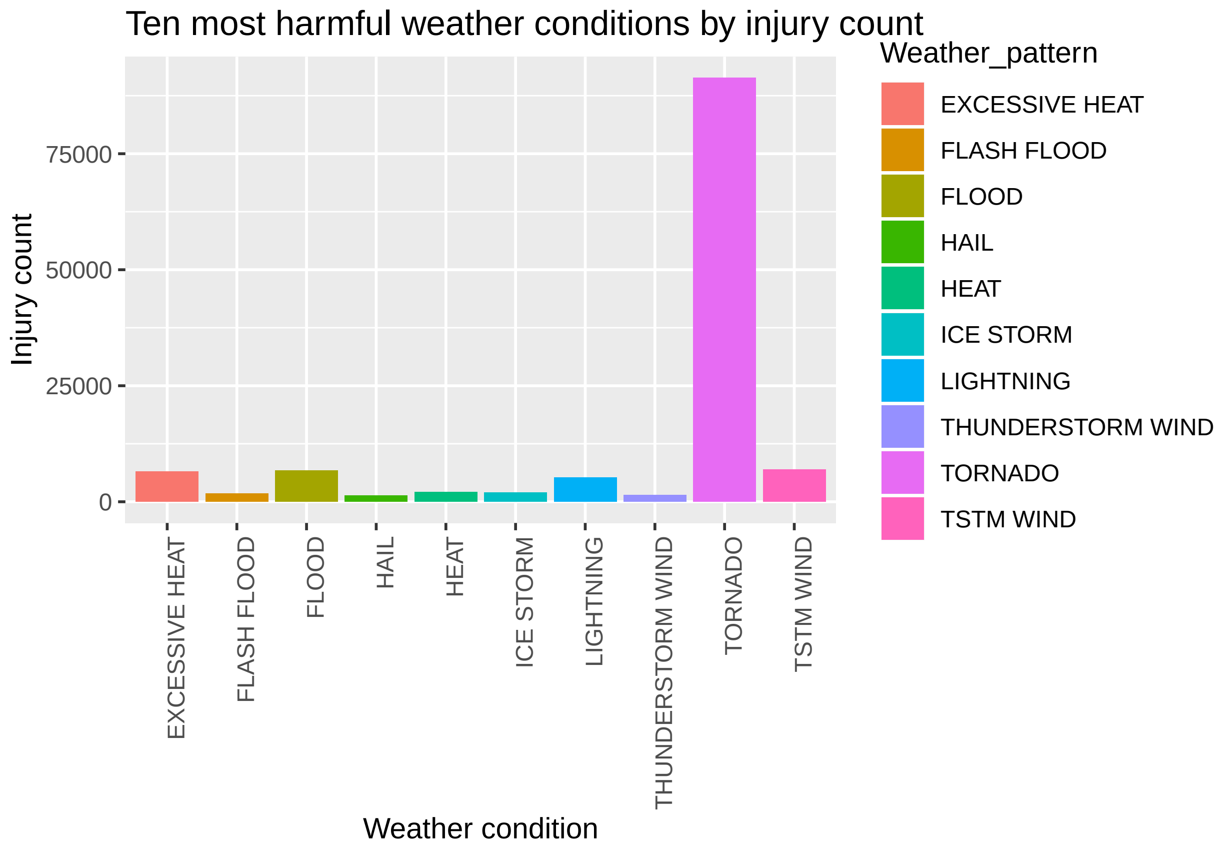
```
most_injuries <- ggplot(top_ten_i, aes(Weather_pattern,injury_count)) +geom_col(aes(fill=Weather_pattern

most_fatalities<- ggplot(top_ten_f, aes(Weather_pattern,fatality_count)) +geom_col(aes(fill=Weather_pat

most_fatalities
```




```
most_injuries
```



2) Across the United States, which types of events have the greatest economic consequences?

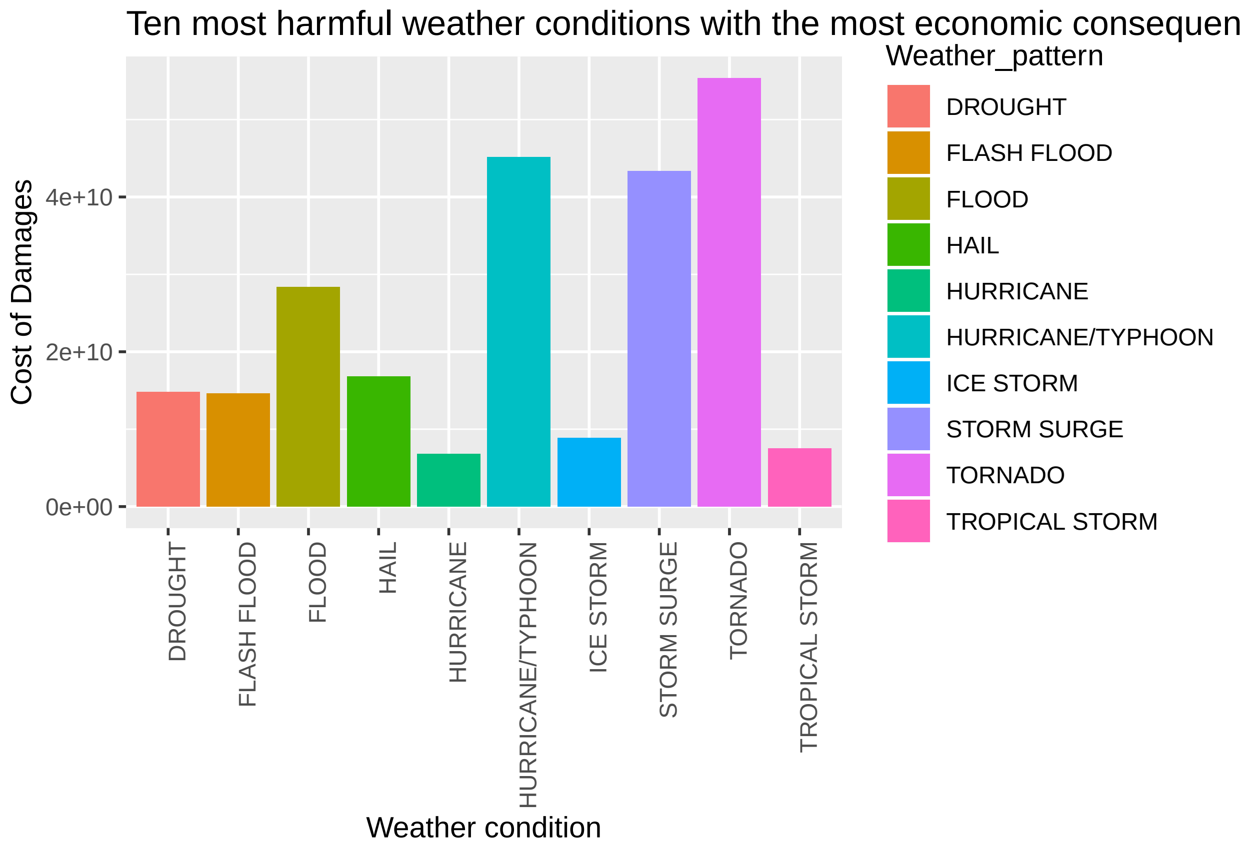
Using the total damage data that we have already calculated and introduced as a seperate column. We now subset the data by weather pattern again and calulate the sum of economic damage for the top 10 weather patterns

```
Economic_damage <- Stormdata_subset %>% group_by(EVTYPE)%>%
  summarise(Economic_dmg =sum(Totaldam,na.rm = TRUE)) %>% as.data.frame() %>% rename(Weather_pattern=EV
top_ten_E <- head(subset(Economic_damage[order(Economic_damage$Economic_dmg,decreasing = TRUE),]),10)
top_ten_E
```

```
##       Weather_pattern Economic_dmg
## 834           TORNADO  55365052590
## 411 HURRICANE/TYPHOON  45173417800
## 670       STORM SURGE  43321941000
## 170             FLOOD  28348588200
## 244              HAIL  16797832730
## 95            DROUGHT  14785940000
## 153       FLASH FLOOD  14635147600
## 427         ICE STORM   8880537310
## 848    TROPICAL STORM   7518856150
## 402         HURRICANE   6800951010
```

```
most_Economics<- ggplot(top_ten_E, aes(Weather_pattern,Economic_dmg)) +geom_col(aes(fill=Weather_patter

most_Economics
```



##Conclusions The report concludes three plots and tables with information about the top 10 most destructive weather patterns with most injuries, fatalities, and economic consequences.