Across the United States, which types of events are most harmful with respect to population health and have the greatest economic consequences?

7/04/2021

Synopsis

In this document, I analyzed the storm data from U.S. National Oceanic and Atmospheric Administration's (NOAA) with the aim of finding out the natural events that are most harmful to the population health and the natural events that have the greatest economic consequences. After I analyzed the data, I found out that Tornados have the biggest impact on the population health, while Floods have the biggest financial damage.

Session info

- For a better reproducibility, here is the session info:

```
sessionInfo()
```

```
## R version 4.0.4 (2021-02-15)
## 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
##
## loaded via a namespace (and not attached):
##  [1] compiler_4.0.4    magrittr_2.0.1    tools_4.0.4
htmltools_0.5.1.1
##  [5] yaml_2.2.1        stringi_1.5.3     rmarkdown_2.9     knitr_1.31
##  [9] stringr_1.4.0     xfun_0.24         digest_0.6.27     rlang_0.4.10
## [13] evaluate_0.14
```

Data Processing

Download, Loading and Processing the data

- Download and Load Packages into RStudio

```
suppressMessages(suppressWarnings(library(dplyr)))
suppressMessages(suppressWarnings(library(ggplot2)))
suppressMessages(suppressWarnings(library(plyr)))
```

- Download data into the computer

```
download.file("https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2", "Storm-Data.zip")
```

- Loading the data into RStudio

```
storm.data <- read.csv("Storm-Data.zip")
head(storm.data)
```

```
##   STATE__           BGN_DATE BGN_TIME TIME_ZONE COUNTY COUNTYNAME STATE
EVTYPE
## 1       1  4/18/1950 0:00:00     0130       CST     97     MOBILE    AL
TORNADO
## 2       1  4/18/1950 0:00:00     0145       CST      3    BALDWIN    AL
TORNADO
## 3       1  2/20/1951 0:00:00     1600       CST     57    FAYETTE    AL
TORNADO
## 4       1   6/8/1951 0:00:00     0900       CST     89    MADISON    AL
TORNADO
## 5       1 11/15/1951 0:00:00     1500       CST     43    CULLMAN    AL
TORNADO
## 6       1 11/15/1951 0:00:00     2000       CST     77 LAUDERDALE    AL
TORNADO
##   BGN_RANGE BGN_AZI BGN_LOCATI END_DATE END_TIME COUNTY_END COUNTYENDN
## 1         0                                               0         NA
## 2         0                                               0         NA
## 3         0                                               0         NA
## 4         0                                               0         NA
## 5         0                                               0         NA
## 6         0                                               0         NA
##   END_RANGE END_AZI END_LOCATI LENGTH WIDTH F MAG FATALITIES INJURIES
PROPDMG
## 1         0                      14.0   100 3   0          0       15
25.0
## 2         0                       2.0   150 2   0          0        0
2.5
## 3         0                       0.1   123 2   0          0        2
25.0
```

```
## 4         0                    0.0   100 2   0          0        2
2.5
## 5         0                    0.0   150 2   0          0        2
2.5
## 6         0                    1.5   177 2   0          0        6
2.5
##   PROPDMGEXP CROPDMG CROPDMGEXP WFO STATEOFFIC ZONENAMES LATITUDE
LONGITUDE
## 1          K       0                                       3040
8812
## 2          K       0                                       3042
8755
## 3          K       0                                       3340
8742
## 4          K       0                                       3458
8626
## 5          K       0                                       3412
8642
## 6          K       0                                       3450
8748
##   LATITUDE_E LONGITUDE_ REMARKS REFNUM
## 1       3051       8806              1
## 2          0          0              2
## 3          0          0              3
## 4          0          0              4
## 5          0          0              5
## 6          0          0              6
```

- Processing the data *In order to answer the question regarding the population health, we only will need three collumns of the database:*

```
population_health_data <- storm.data %>% select (EVTYPE, INJURIES,
FATALITIES)
head(population_health_data)
```

```
##    EVTYPE INJURIES FATALITIES
## 1 TORNADO       15          0
## 2 TORNADO        0          0
## 3 TORNADO        2          0
## 4 TORNADO        2          0
## 5 TORNADO        2          0
## 6 TORNADO        6          0
```

- Break it down the data for the type of events

```
summary(unique(population_health_data$EVTYPE))
```

```
##    Length     Class      Mode
##       985 character character
```

There are 985 unique storm events. Let's create a new collumn called health.risk with the sum of injuries and fatalities:

```
population_health_data <- ddply(population_health_data,.(EVTYPE), summarize,
injuries= sum(INJURIES), fatalities = sum(FATALITIES), health.risk =
sum(INJURIES) + sum(FATALITIES))
head(population_health_data)
```

```
##                  EVTYPE injuries fatalities health.risk
## 1    HIGH SURF ADVISORY        0          0           0
## 2         COASTAL FLOOD        0          0           0
## 3           FLASH FLOOD        0          0           0
## 4             LIGHTNING        0          0           0
## 5             TSTM WIND        0          0           0
## 6       TSTM WIND (G45)        0          0           0
```

Delete rows that have zero health.risk:

```
population_health_data <-
population_health_data[!(population_health_data$health.risk == 0),]
head(population_health_data)
```

```
##          EVTYPE injuries fatalities health.risk
## 18     AVALANCE        0          1           1
## 19    AVALANCHE      170        224         394
## 29    BLACK ICE       24          1          25
## 30     BLIZZARD      805        101         906
## 42 blowing snow        1          1           2
## 44 BLOWING SNOW       13          1          14
```

Find 10 largest health related risks:

```
biggest_health_risk <- population_health_data[with(population_health_data,
order(-health.risk)),]
biggest_health_risk <- biggest_health_risk[1:10,]
head(biggest_health_risk)
```

```
##             EVTYPE injuries fatalities health.risk
## 834        TORNADO    91346       5633       96979
## 130 EXCESSIVE HEAT     6525       1903        8428
## 856      TSTM WIND     6957        504        7461
## 170          FLOOD     6789        470        7259
## 464      LIGHTNING     5230        816        6046
## 275           HEAT     2100        937        3037
```

- Transform and factor data to be used in the plot

```
biggest_health_risk <- transform(biggest_health_risk,
                                 EVTYPE = factor(EVTYPE),
                                 injuries = factor(injuries),
                                 fatalities = factor(fatalities),
```

```
                                health.risk = factor(health.risk)
)
```

In order to answer the question regarding the financial damage, we will only need 5 collumns:

```
financial_damage <- storm.data %>% select (EVTYPE, PROPDMG, PROPDMGEXP,
CROPDMG, CROPDMGEXP)
head(financial_damage)

##    EVTYPE PROPDMG PROPDMGEXP CROPDMG CROPDMGEXP
## 1 TORNADO    25.0          K       0
## 2 TORNADO     2.5          K       0
## 3 TORNADO    25.0          K       0
## 4 TORNADO     2.5          K       0
## 5 TORNADO     2.5          K       0
## 6 TORNADO     2.5          K       0
```

- Now, we are going to convert the letters to actual numbers: K: Thousands (1000) M: Millions (1000000) B: Billions (1000000000)

```
financial_damage[financial_damage$PROPDMGEXP == "K",]$PROPDMGEXP = 1000
financial_damage[financial_damage$PROPDMGEXP == "M",]$PROPDMGEXP = 1000000
financial_damage[financial_damage$PROPDMGEXP == "B",]$PROPDMGEXP = 1000000000
financial_damage[financial_damage$CROPDMGEXP == "K",]$CROPDMGEXP = 1000
financial_damage[financial_damage$CROPDMGEXP == "M",]$CROPDMGEXP = 1000000
financial_damage[financial_damage$CROPDMGEXP == "B",]$CROPDMGEXP = 1000000000
financial_damage[financial_damage$CROPDMGEXP == "",]$CROPDMGEXP = 0
#financial_damage[financial_damage$PROPDMG == "",]$PROPDMG = 0
financial_damage[financial_damage$PROPDMGEXP == "",]$PROPDMGEXP = 0
#financial_damage[financial_damage$CROPDMG == "",]$CROPDMG = 0
```

- Check Classes of data

```
sapply(financial_damage, class)

##      EVTYPE     PROPDMG  PROPDMGEXP     CROPDMG  CROPDMGEXP
## "character"   "numeric" "character"   "numeric" "character"

financial_damage$PROPDMGEXP <-
as.numeric(as.character(financial_damage$PROPDMGEXP))

## Warning: NAs introduced by coercion

financial_damage$CROPDMGEXP <-
as.numeric(as.character(financial_damage$CROPDMGEXP))

## Warning: NAs introduced by coercion
```

- Create new collumn called Total Financial Damage:

```
financial_damage$fin.damage <- (financial_damage$PROPDMG *
financial_damage$PROPDMGEXP) + (financial_damage$CROPDMG *
```

```
financial_damage$CROPDMGEXP)
head(financial_damage)

##    EVTYPE PROPDMG PROPDMGEXP CROPDMG CROPDMGEXP fin.damage
## 1 TORNADO    25.0       1000       0          0      25000
## 2 TORNADO     2.5       1000       0          0       2500
## 3 TORNADO    25.0       1000       0          0      25000
## 4 TORNADO     2.5       1000       0          0       2500
## 5 TORNADO     2.5       1000       0          0       2500
## 6 TORNADO     2.5       1000       0          0       2500
```

Find 10 largest financial damage:

```
biggest_financial_damage <- financial_damage[with(financial_damage, order(-
fin.damage)),]
biggest_financial_damage <- biggest_financial_damage[1:10,]
head(biggest_financial_damage)

##                   EVTYPE PROPDMG PROPDMGEXP CROPDMG CROPDMGEXP
fin.damage
## 605953             FLOOD  115.00      1e+09    32.5      1e+06
115032500000
## 577676       STORM SURGE   31.30      1e+09     0.0      0e+00
31300000000
## 577675 HURRICANE/TYPHOON   16.93      1e+09     0.0      0e+00
16930000000
## 581535       STORM SURGE   11.26      1e+09     0.0      0e+00
11260000000
## 198389       RIVER FLOOD    5.00      1e+09     5.0      1e+09
10000000000
## 569308 HURRICANE/TYPHOON   10.00      1e+09     0.0      0e+00
10000000000
```

A quick look at the table and we can see that flood has the biggest financial damage.

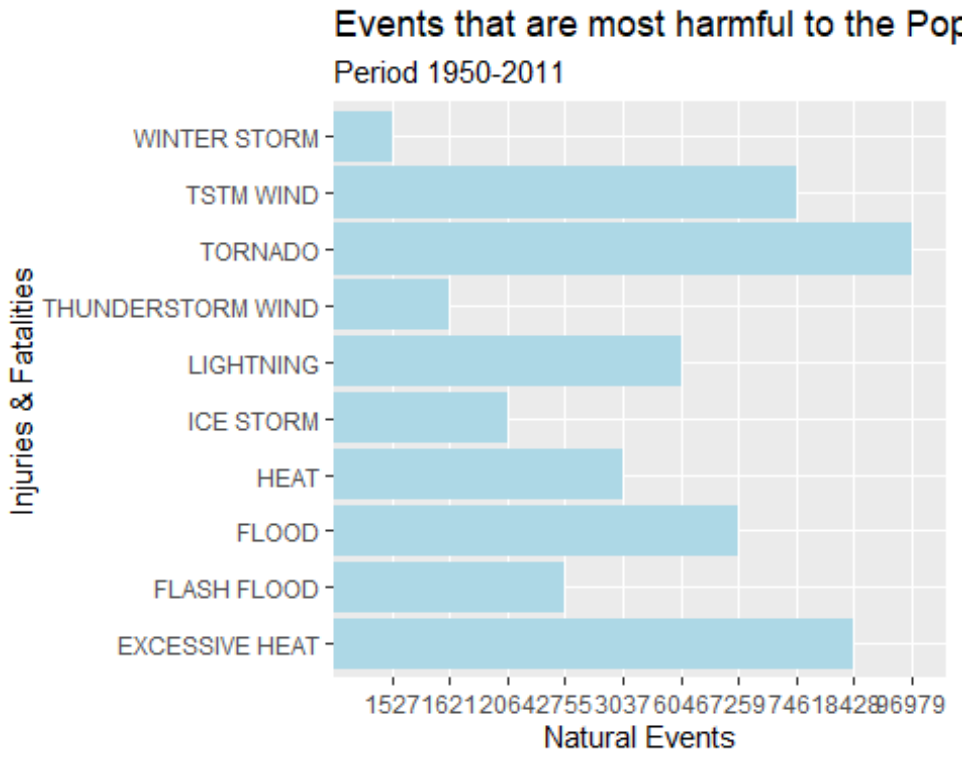
Results

Which types of events are most harmful with respect to population health? - Create Plot for the Population Health data

```
pop_health_plot <- ggplot(data = biggest_health_risk, aes(x = EVTYPE, y =
health.risk)) +
geom_bar(fill = "lightblue", stat = "identity", position = "dodge")
pop_health_plot <- pop_health_plot + coord_flip()
```

- Add info to the plot

```
pop_health_plot + labs(
        title = "Events that are most harmful to the Population Health",
        subtitle = "Period 1950-2011",
        x = "Injuries & Fatalities",
        y = "Natural Events")
```



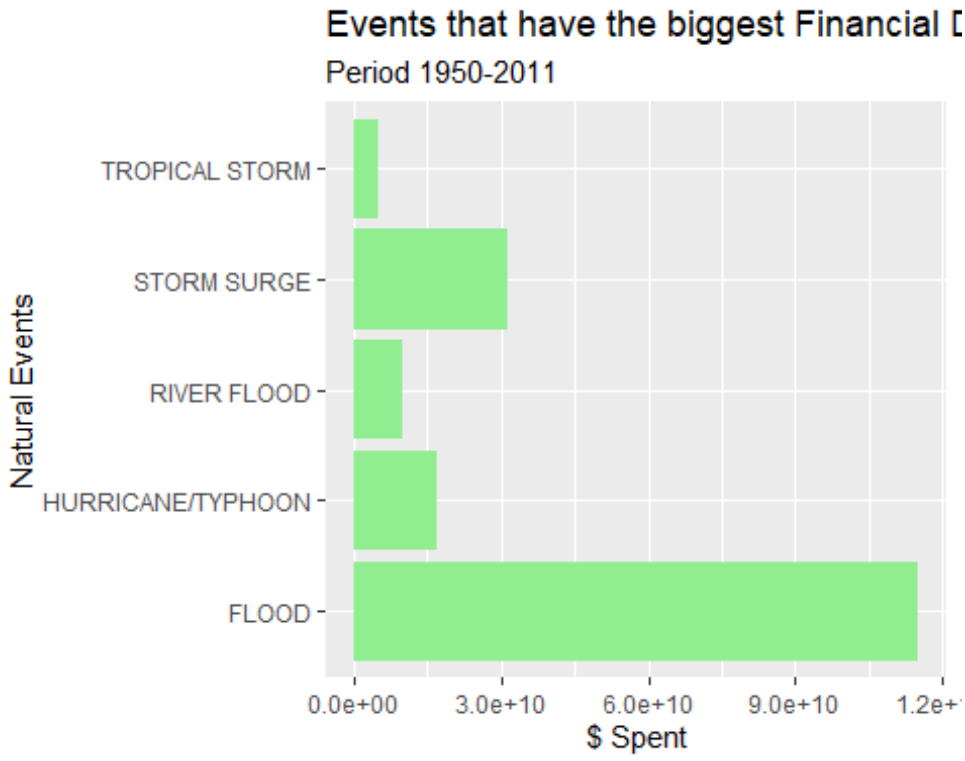
- As we can see in the graph, Tornados are the most harmful to the health.

Which types of events have the biggest financial damage? - Create Plot for the Financial Damage data

```
pop_fin_plot <- ggplot(data = biggest_financial_damage, aes(x = EVTYPE, y =
fin.damage)) +
geom_bar(fill = "lightgreen", stat = "identity", position = "dodge")
pop_fin_plot <- pop_fin_plot + coord_flip()
```

- Add info to the plot

```
pop_fin_plot + labs(
        title = "Events that have the biggest Financial Damage",
        subtitle = "Period 1950-2011",
        x = "Natural Events",
        y = "$ Spent")
```



The graph confirms that flood has the biggest financial damage.