US Storm Event Impact Analysis

Rohith Shankar

6/24/2020

Synopsis

This is analysis done on the impact of Storm Events on the US economy as well as the Popukation Health. The data for this analysis is obtained from the National Weather Service website. The data covers the period between 1950 and 2011. The analysis addresses the following two questions.

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

Data Processing

```
#Read the storm data csv and understand it
setwd(".\\data")
StormData <- read.csv(".\\repdata_data_StormData.csv", sep=",", header=TRUE)

dim(StormData)
```

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

```
names(StormData)
```

```
##  [1] "STATE__"    "BGN_DATE"   "BGN_TIME"   "TIME_ZONE"  "COUNTY"
##  [6] "COUNTYNAME" "STATE"      "EVTYPE"     "BGN_RANGE"  "BGN_AZI"
## [11] "BGN_LOCATI" "END_DATE"   "END_TIME"   "COUNTY_END" "COUNTYENDN"
## [16] "END_RANGE"  "END_AZI"    "END_LOCATI" "LENGTH"     "WIDTH"
## [21] "F"          "MAG"        "FATALITIES" "INJURIES"   "PROPDMG"
## [26] "PROPDMGEXP" "CROPDMG"    "CROPDMGEXP" "WFO"        "STATEOFFIC"
## [31] "ZONENAMES"  "LATITUDE"   "LONGITUDE"  "LATITUDE_E" "LONGITUDE_"
## [36] "REMARKS"    "REFNUM"
```

```
str(StormData)
```

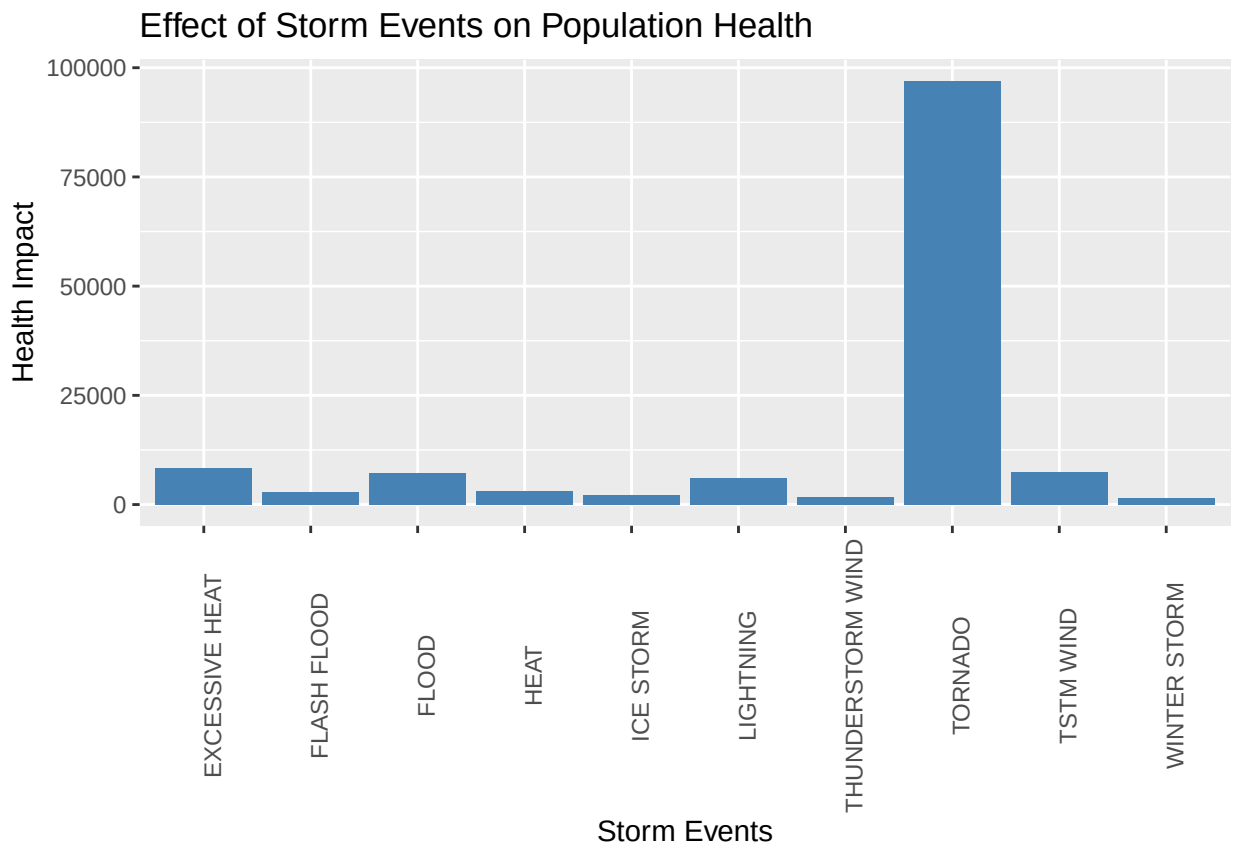
```
## 'data.frame':    902297 obs. of  37 variables:
##  $ STATE__   : num  1 1 1 1 1 1 1 1 1 1 ...
##  $ BGN_DATE  : Factor w/ 16335 levels "1/1/1966 0:00:00",..: 6523 6523 4242 11116 2224 2224 2260 383
##  $ BGN_TIME  : Factor w/ 3608 levels "00:00:00 AM",..: 272 287 2705 1683 2584 3186 242 1683 3186 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 1
##  $ 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/ 436774 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 ...
```

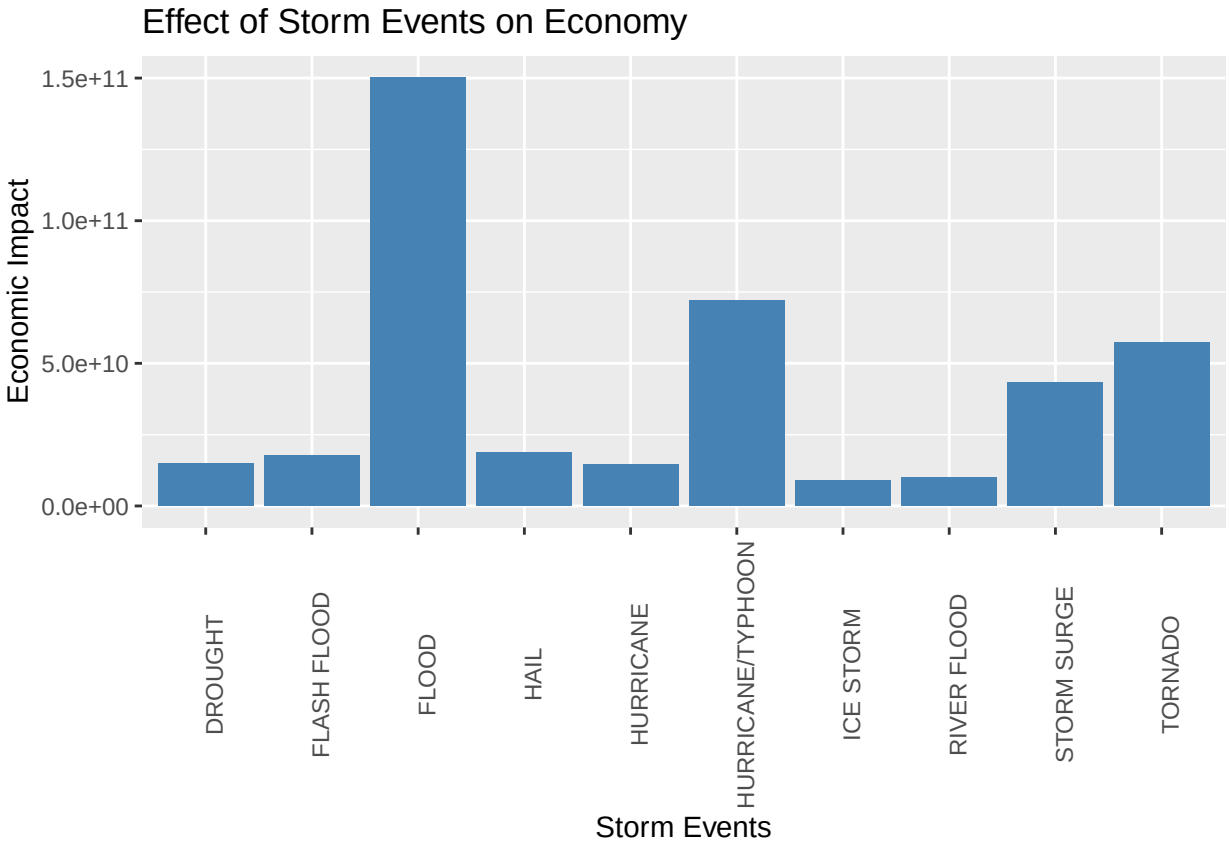
```
library(dplyr)

# Choose only the relevant columns
StormData = select(StormData, EVTYPE, FATALITIES, INJURIES, PROPDMG, PROPDMGEXP, CROPDMG, CROPDMGEXP)

# Transform the Property Damage and Crop Damage to reflect actual numerical value
# Only with this transformation can we actually comapre the impact and plot them
StormData$PROPDMG <- case_when(StormData$PROPDMGEXP=='H' ~ StormData$PROPDMG *10^2
                               ,StormData$PROPDMGEXP=='K' ~ StormData$PROPDMG *10^3
                               ,StormData$PROPDMGEXP=='M' ~ StormData$PROPDMG *10^6
                               ,StormData$PROPDMGEXP=='B' ~ StormData$PROPDMG *10^9
                               ,TRUE ~ StormData$PROPDMG)

StormData$CROPDMG <- case_when(StormData$CROPDMGEXP=='H' ~ StormData$CROPDMG *10^2
                               ,StormData$CROPDMGEXP=='K' ~ StormData$CROPDMG *10^3
                               ,StormData$CROPDMGEXP=='M' ~ StormData$CROPDMG *10^6
                               ,StormData$CROPDMGEXP=='B' ~ StormData$CROPDMG *10^9
                               ,TRUE ~ StormData$CROPDMG)
```

```r
# Get the total crop damage and property by storm event
EvtDamage <- summarize(group_by(StormData, EVTYPE), total_propdmg=sum(PROPDMG),total_cropdmg=sum(CROPDM

# Determine the total economic damage
EvtDamage$ECODMG <- EvtDamage$total_propdmg+EvtDamage$total_cropdmg

# Choose the top 10 events by economic damage
EcoDamage_Final<-head(arrange(EvtDamage,desc(ECODMG)), n=10)

# Get the total fatalities and injuries by storm event
PopDamage <- summarize(group_by(StormData, EVTYPE), total_fatalities=sum(FATALITIES)
                       , total_injuries=sum(INJURIES))

#Determine the total health damage by storm event
PopDamage$TotalPopDmg <-PopDamage$total_fatalities + PopDamage$total_injuries

# Choose the top 10 events by population health damage
PopDamage_Final <- head(arrange(PopDamage,desc(TotalPopDmg)), n=10)
```

```r
library(ggplot2)

# Bar plot of the top 10 storm events by health impact
ggplot(PopDamage_Final, aes(x=EVTYPE, y=TotalPopDmg))+
geom_bar(stat="identity", fill="steelblue")+
scale_x_discrete(name="Storm Events") + scale_y_continuous(name="Health Impact")+
theme(axis.text.x = element_text(angle = 90)) +
theme(axis.text.y = element_text(angle=0))+
labs(title = "Effect of Storm Events on Population Health")
```




```
# Bar plot of the top 10 storm events by economic impact
ggplot(EcoDamage_Final, aes(x=EVTYPE, y=ECODMG))+
geom_bar(stat="identity", fill="steelblue")+
scale_x_discrete(name="Storm Events") + scale_y_continuous(name="Economic Impact")+
theme(axis.text.x = element_text(angle = 90)) +
theme(axis.text.y = element_text(angle=0))+
labs(title = "Effect of Storm Events on Economy")
```



Results

1. It is clear from the above bar plot the that tornadoes have the most impact on population health causing injuries and fatalities far above every other event.

2. Flood seems to cause the maximum economic impact although Hurricanes and Tornadoes also contribute significantly towards property and crop damage.