Economic and Health Impacts of Disasters

Deepak Gogoi

7/26/2020

Synopsis

This report analyzes the effect of different type of disasters on health and economy. The data set contains more than 900,000 disaster events and their impacts. The data were used to show the most harmful disasters (events) for both economic situation and population health. The process used for analysis includes cleaning data and performing a pareto chart to obtain the fewest number of events responsible for most of the economic and health consequences. For the health impact, both fatalities and injuries were aggregated to get total casualties and all events with no casualties were excluded. Later on, the events which had less than 1% of the total casualties were also removed to get the pareto chart. For the economic consequences, the crop and property damages were cleaned, combined and then added together. The same procedure was used to eliminate events with no economic losses and with percentage less than 2% of the total.

Data Processing

The steps used for data processing can be described as follows:

First, Loading the data and the packages used

```
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
```

```
library(ggplot2)
storm <- read.csv("repdata_data_StormData.csv.bz2")
```

Economic Consequences:

1- The events with no fatalities or injuries were excluded (assumed that each event to have a significant effect has to have both). The total of fatalities and injuries per event was calculated, and then ordered and summed to get the total casualties.

```
st<- subset(storm, storm$FATALITIES >0 )
st<- subset(st, st$INJURIES >0 )
byevent<- st %>% group_by(EVTYPE) %>% summarise(fat= sum(FATALITIES), inj=
sum(INJURIES))
```

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

```
byevent$cas <- byevent$fat + byevent$inj #total casualties

byevent <- byevent[order(-byevent$cas),] #order the data set

byevent$cascum<- cumsum(byevent$cas) #total cumulative casualties
```

2-The percentage of casualties per event was calculated, and the cumulative percentage. The events with less than 1% of total casualties were removed.

```
byevent <- mutate(byevent, cas_pct=cas/sum(cas)) #percentage of casualties per event

byevent$cascumpct<- cumsum(byevent$cas_pct) #cumulative percentage of casualties per event

major<- subset(byevent,byevent$cas_pct>0.01) #eliminate percentages <1%
```

Economic Consequences:

In this part, the economic consequences were quantified as follows: the property damage “PROPDMG” and the crop damage “CROPDMG” were aggregated. However, since the units are not the same for each observation as shown in the link provided in the project statement in National Weather Service Storm Data Documentation, they had to be unified. The first step was to find out which ones are in hundreds, million, thousands, etc. which are presented in variables “PROPDMGEXP” and “CROPDMGEXP”. Upon inspecting these variables, they turned out to have unexplained symbols. so to deal with that, any unexplained value was taken to have a multiplier of zero to eliminate the observation. Other observations were multiplied by a multiplier to get the correct value, e.g. 25k –> 25*1000 and so on.

1- The two variables for the unit of damage cost were inspected to see if they have missing values

```
sum(is.na(storm$PROPDMGEXP))

## [1] 0

sum(is.na(storm$CROPDMGEXP))

## [1] 0
```

2- Obtaining the unique values for the units in variables “PROPDMGEXP” and “CROPDMGEXP”, and creating dictionaries for these values by merging them to a especially created dictionary.It should be noted that the values of property and crop damage were taken face value ,i.e., with no multipliers for the observations that had empty values in “PROPDMGEXP” and “CROPDMGEXP”.

```
dict_p<- as.data.frame(unique(storm$PROPDMGEXP)) #unique values of the "PROPDMGEXP" column

names(dict_p) <- c("PROPDMGEXP")

dict<- as.data.frame(cbind(c("B","b","M","m","K","k","H","h",""),c(1000000000,1000000000,1000000,1000000,1000,1000,100,100,1))) #custom made dictionary to unify the units

names(dict) <- c("PROPDMGEXP","PROPmultiplier")

dict_p<- merge(dict_p,dict,by="PROPDMGEXP" ,all.x = TRUE) #merging the dictionary with the unique value to get corresponding value for each input of property damage

dict_c<- as.data.frame(unique(storm$CROPDMGEXP))  #unique values of the "CROPDMGEXP" column

names(dict_c) <- c("CROPDMGEXP")

names(dict) <- c("CROPDMGEXP","CROPmultiplier")
```

```
dict_c<- merge(dict_c,dict,by="CROPDMGEXP" ,all.x = TRUE) #merging the dict
ionary with the unique value to get corresponding value for each input of c
rop damage
```

3- The data sets were merged with both property and crop damage dictionaries. The observations with strange unexplained symbols were eliminated, then economic losses were aggregated (crop + property damages)

```
st<- merge(storm,dict_p,by="PROPDMGEXP",all.x = TRUE) #merging the property
 damage dictionary

st<- merge(st,dict_c,by="CROPDMGEXP",all.x = TRUE) #merging the crop damage
 dictionary

#changing the missing values (unexplained symbols) into zeros to eliminate
the observations

st$PROPmultiplier[is.na(st$PROPmultiplier)] <- 0

st$CROPmultiplier[is.na(st$CROPmultiplier)] <- 0

#changing the new added columns into numeric

st$PROPmultiplier<- as.numeric(st$PROPmultiplier)

st$CROPmultiplier<- as.numeric(st$CROPmultiplier)

#aggregating the economic losses

st$econ <- st$PROPmultiplier*st$PROPDMG + st$CROPmultiplier*st$CROPDMG
```

4- The events with no economic losses were excluded. Then the total economic losses by event type were calculated. Using these values, the percentages were calculated and the events with losses of 2% of the total were excluded

```
econ <- subset(st,st$econ>0) #exclude events with no economic losses

#total economic losses by event type

byevent_e <- econ %>% group_by(EVTYPE) %>% summarise(totval= sum(econ))
```

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

```
byevent_e <- byevent_e[order(-byevent_e$totval),] #ordering the events by l
osses (descending)

byevent_e$econcum<- cumsum(byevent_e$totval) #cumulative sum of economic lo
sses

byevent_e <- mutate(byevent_e, econ_pct=totval/sum(totval)) #percentage of
losses from total

byevent_e$econcumpct<- cumsum(byevent_e$econ_pct) #cumulative percentage of
 all events

#excluding events with less than 2% of the total losses

majorecon<- subset(byevent_e,byevent_e$econ_pct>0.02)
```

Results

Population Health Impacts

The types of events that are most harmful with respect to population health according to the proposed analysis:

```
print(major)
```

```
## # A tibble: 7 x 7
##   EVTYPE              fat   inj   cas cascum cas_pct cascumpct
##   <chr>             <dbl> <dbl> <dbl>  <dbl>   <dbl>     <dbl>
## 1 TORNADO            5227 60187 65414  65414  0.756      0.756
## 2 EXCESSIVE HEAT      402  4791  5193  70607  0.0600     0.816
## 3 FLOOD               104  2679  2783  73390  0.0321     0.848
## 4 ICE STORM            35  1720  1755  75145  0.0203     0.868
## 5 HEAT                 73  1420  1493  76638  0.0172     0.885
## 6 HURRICANE/TYPHOON    32  1219  1251  77889  0.0144     0.900
## 7 LIGHTNING           283   649   932  78821  0.0108     0.910
```

```
print(paste0("These ", nrow(major)," event types are responsible for ", rou
nd(sum(major$cas_pct)*100,2), " % of the total casualties casued by disaste
r in the US"))
```

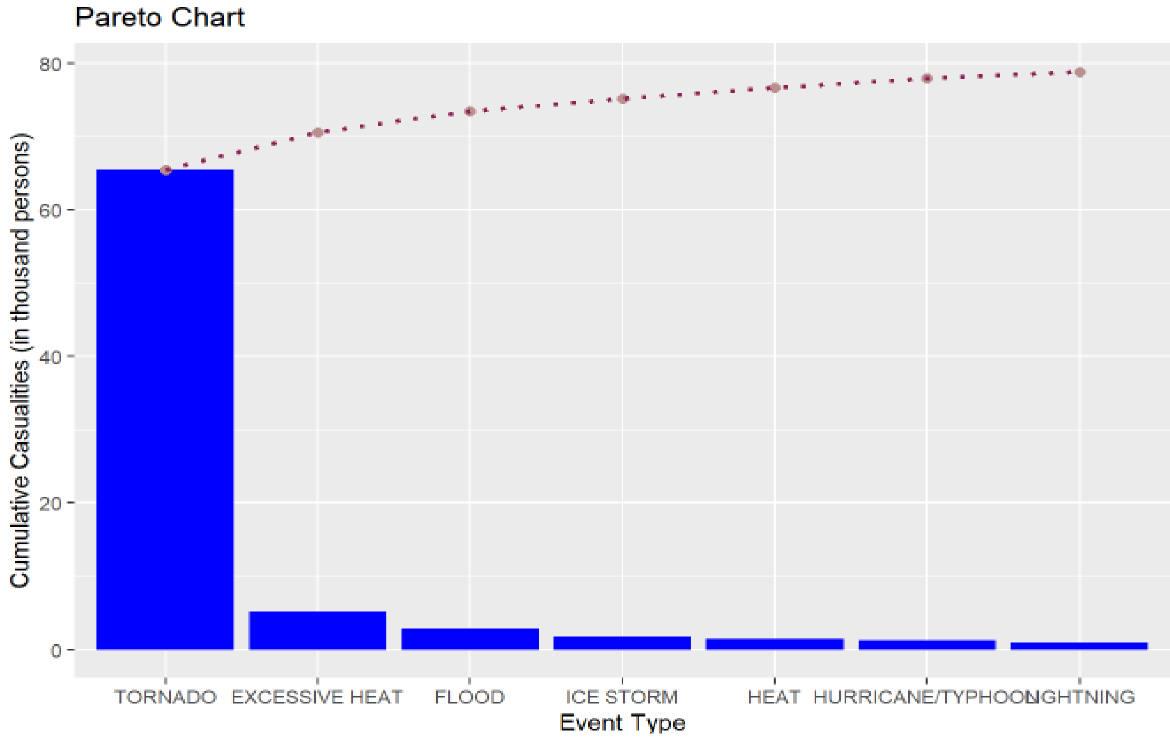
```
## [1] "These 7 event types are responsible for 91.04 % of the total casual
ties casued by disaster in the US"
```

The pareto chart of the cumulative casualties caused by major disasters:

```
g<- ggplot (major,aes(x=EVTYPE))

g2<- g+ geom_bar(aes(y=cas/1000),fill='blue',stat = "identity") + scale_x_d
iscrete(limits=major$EVTYPE)

g2+ geom_point(aes(y=cascum/1000),color='rosybrown',pch=16,size=2)+geom_pat
h(aes(y=cascum/1000, group=1),color="violetred4",lty=3,size=0.9)+labs(title
 = "Pareto Chart", y="Cumulative Casualities (in thousand persons)", x="Eve
nt Type")
```



Economic Consequences

The types of events that are most harmful with respect to population health according to the proposed analysis:

```
print(majorecon)
```

```
## # A tibble: 9 x 5
##   EVTYPE                  totval      econcum econ_pct econcumpct
##   <chr>                    <dbl>        <dbl>    <dbl>      <dbl>
## 1 FLOOD             150319678257 150319678257   0.316       0.316
## 2 HURRICANE/TYPHOON  71913712800 222233391057   0.151       0.466
## 3 TORNADO            57352113593 279585504650   0.120       0.587
## 4 STORM SURGE        43323541000 322909045650   0.0909      0.678
## 5 HAIL               18758221730 341667267380   0.0394      0.717
## 6 FLASH FLOOD        17562128817 359229396197   0.0369      0.754
## 7 DROUGHT            15018672000 374248068197   0.0315      0.786
## 8 HURRICANE          14610229010 388858297207   0.0307      0.816
## 9 RIVER FLOOD        10148404500 399006701707   0.0213      0.838
```

```
print(paste0("These ", nrow(majorecon)," event types are responsible for ", round(sum(majorecon$econ_pct)*100,2), " % of the total economic losses casued by disaster in the US"))
```

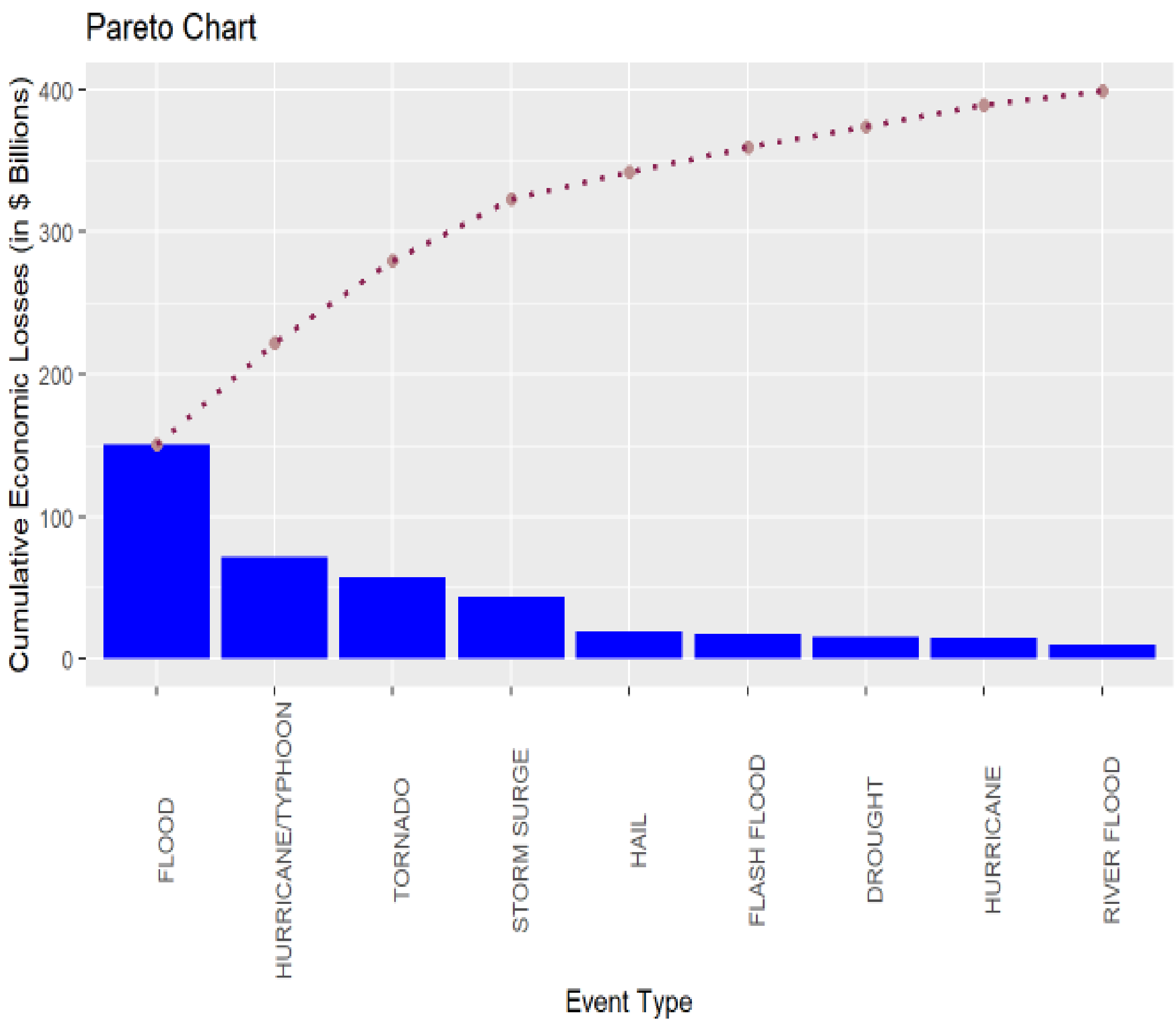
```
## [1] "These 9 event types are responsible for 83.75 % of the total economic losses casued by disaster in the US"
```

The pareto chart of the cumulative economic losses caused by major disasters:

```
e<- ggplot (majorecon,aes(x=EVTYPE))

e2<- e+ geom_bar(aes(y=totval/1000000000),fill='blue',stat = "identity") + scale_x_discrete(limits=majorecon$EVTYPE)

e2+ geom_point(aes(y=econcum/1000000000),color='rosybrown',pch=16,size=2)+geom_path(aes(y=econcum/1000000000, group=1),color="violetred4",lty=3,size=0.9)+labs(title = "Pareto Chart", y="Cumulative Economic Losses (in $ Billions)", x="Event Type")+theme(axis.text.x = element_text(angle=90)
```



Pareto Chart
Cumulative Economic Losses (in $ Billions)
400
300
200
100
0
FLOOD
HURRICANE/TYPHOON
TORNADO
STORM SURGE
HAIL
FLASH FLOOD
DROUGHT
HURRICANE
RIVER FLOOD
Event Type