Analysis of Economic and Population Health Impact of Historical Storm Events

Atharva Ramgirkar

07/13/2020

Synopsis

Dataset overview

The National Oceanic and Atmospheric Administration (NOAA) maintains a public database for storm event. The data contains the type of storm event, details like location, date, estimates for damage to property as well as the number of human victims of the storm.

Goal

Our goal in this report will be to answer two important questions:- 1. Across the United States, which types of events (as indicated in 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 Methodology

For the first question, To quantify the harm to population health, we find the total number of fatalities and the total number of injuries and check which weather type(*evtype*) was responsible for majority of them. We plot the result to make it easy to interpret.

For the second question, We categorize the property damage and crop damage. Then we plot the net damage in millions of USD v/s the weather type(*evtype*) which was responsible for it.

Data Processing

Load the necessary libraries

```
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(plyr) # for count & aggregate method
```

```
## ------------------------------------------------------------------------------
```

```
## You have loaded plyr after dplyr - this is likely to cause problems.
## If you need functions from both plyr and dplyr, please load plyr first, then dplyr:
## library(plyr); library(dplyr)
```

```
## ------------------------------------------------------------------------------
```

```
##
## Attaching package: 'plyr'
```

```
## The following objects are masked from 'package:dplyr':
##
##     arrange, count, desc, failwith, id, mutate, rename, summarise,
##     summarize
```

```
library(ggplot2) # for plots
```

```
## Warning: package 'ggplot2' was built under R version 4.0.2
```

```
library(knitr)
library(scales) # for plot scaling
options(scipen = 1) # Turn off scientific notations for numbers
library(lubridate)
```

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

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

```
library("cowplot")
```

```
## Warning: package 'cowplot' was built under R version 4.0.2
```

```
##
## ************************************************************
```

```
## Note: As of version 1.0.0, cowplot does not change the
```

```
##   default ggplot2 theme anymore. To recover the previous
```

```
##   behavior, execute:
##   theme_set(theme_cowplot())

## ******************************************************

##
## Attaching package: 'cowplot'

## The following object is masked from 'package:lubridate':
##
##     stamp
```

```
library(reshape2)
```

```
knitr::opts_chunk$set(echo = TRUE, warning = FALSE, cache = TRUE)
```

set the default options for the RMarkdown so we don't need to type these in every codeblock

```
setwd("F:/JohnHopkindDS/Course5-ReproducibleResearch/Reproducible-Research-Data-Science/Week4/Graded Ass
```

Set an appropriate working directory

Step 1: Reading the Data from the given URL and downloading it to the working directory

```
if (!"FStormData.csv.bz2" %in% dir("./"))
  {

      download.file(url ="https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2",
                    destfile = "FStormData.csv.bz2")
  }
```

Read the csv file and store it in a variable

```
if(!"storm.data" %in% ls())
  {
storm.data<-read.csv(file = "FStormData.csv.bz2",
                     header = TRUE,
                     sep = ",")
  }
```

modify the names of the columns so they can be easily referenced

1. Convert all the column names to lower case
2. Take only the relevant columns for analysis(*10/48 columns retrieved*)
3. Filter and use the data only after 1996 as it is well documented
4. Make a **cropdmg_total** and **propdmg_total** to store the total crop damage and total property damage
5. Filter out the appropriate **evtype** as given in the document(*there are only 48 event types*)
6. Give the correct value of the damage caused in numeric forms. The exponent is given in (*K,B,M,H*). Make a new column and give numberic values accordingly

```
names(storm.data)<-tolower(names(storm.data))

storm.data.req <- select(storm.data,evtype,bgn_date,fatalities,injuries,propdmg,propdmgexp,cropdmg,crop
mutate(bgn_date = mdy_hms(bgn_date),evtype=tolower(evtype)) %>%
  filter(year(bgn_date) >= 1996)



events.in_doc <-          c("astronomical low tide",
                            "avalanche",
                            "blizzard",
                            "coastal flood",
                            "cold/wind chill",
                            "debris flow",
                            "dense fog",
                            "dense smoke",
                            "drought",
                            "dust devil",
                            "dust storm",
                            "excessive heat",
                            "extreme cold/wind chill",
                            "flash flood",
                            "flood",
                            "frost/freeze",
                            "funnel cloud",
                            "freezing fog",
                            "hail",
                            "heat",
                            "heavy rain",
                            "heavy snow",
                            "high surf",
                            "high wind",
                            "hurricane (typhoon)",
                            "ice storm",
                            "lake-effect snow",
                            "lakeshore flood",
                            "lightning",
                            "marine hail",
                            "marine high wind",
                            "marine strong wind",
                            "marine thunderstorm wind",
                            "rip current",
                            "seiche",
```

```
                              "sleet",
                              "storm surge/tide",
                              "strong wind",
                              "thunderstorm wind",
                              "tornado",
                              "tropical depression",
                              "tropical storm",
                              "tsunami",
                              "volcanic ash",
                              "waterspout",
                              "wildfire",
                              "winter storm",
                              "winter weather")

storm.data.allowed<-subset(storm.data.req,
                           evtype %in% events.in_doc)

storm.data.NOTallowed<-subset(storm.data.req,
                              !evtype %in% events.in_doc)

storm.data_event <- mutate(storm.data.NOTallowed,
  evtype = case_when(
    grepl("marine",evtype,ignore.case = TRUE) ~ "marine thunderstorm wind",
    grepl("tstm|thunde",evtype,ignore.case = TRUE) ~ "thunderstorm wind",
    grepl("fire",evtype,ignore.case = TRUE) ~ "wildfire",
    grepl("coastal",evtype,ignore.case = TRUE) ~ "coastal flood",
    grepl("fld|flood",evtype,ignore.case = TRUE) ~ "flood",
    grepl("winter|snow",evtype,ignore.case = TRUE) ~ "winter weather",
    grepl("cold|chill",evtype,ignore.case = TRUE) ~ "cold/wind chill",
    grepl("fog",evtype,ignore.case = TRUE) ~ "dense fog",
    grepl("hurric|typhoo",evtype,ignore.case = TRUE) ~ "hurricane (typhoon)",
    grepl("warm|heat|hot",evtype,ignore.case = TRUE) ~ "heat",
    grepl("wind",evtype,ignore.case = TRUE) ~ "strong wind",
    grepl("rip ",evtype,ignore.case = TRUE) ~ "rip current",
    grepl("surge",evtype,ignore.case = TRUE) ~ "storm surge/tide",
    grepl("surf",evtype,ignore.case = TRUE) ~ "high surf",
    grepl("blizz",evtype,ignore.case = TRUE) ~ "blizzard",
    grepl("frost|freez|icy|ice", evtype,ignore.case = TRUE) ~ "frost/freeze",
                                    TRUE ~ "other"))

storm.data.event<-rbind(storm.data.allowed,storm.data_event)

storm.data.event<-mutate(storm.data.event,
                         propdmg_total= case_when(
                           propdmgexp=="H"|propdmgexp=="h" ~ propdmg*100,
                           propdmgexp=="K"|propdmgexp=="k" ~ propdmg*1000,
                           propdmgexp=="M"|propdmgexp=="m" ~ propdmg*1000000,
                           propdmgexp=="B"|propdmgexp=="b" ~ propdmg*1000000000,
                           grepl("[0-9]", propdmgexp) ~ propdmg * 10^as.numeric(propdmgexp),
                           TRUE ~ propdmg
                           ),
                         cropdmg_total= case_when(
                           cropdmgexp=="H"|cropdmgexp=="h" ~ cropdmg*100,
```

```
                            cropdmgexp=="K"|cropdmgexp=="k" ~ cropdmg*1000,
                            cropdmgexp=="M"|cropdmgexp=="m" ~ cropdmg*1000000,
                            cropdmgexp=="B"|cropdmgexp=="b" ~ cropdmg*1000000000,
                            grepl("[0-9]", cropdmgexp) ~ cropdmg * 10^as.numeric(cropdmgexp),
                            TRUE ~ cropdmg
                            ))
```

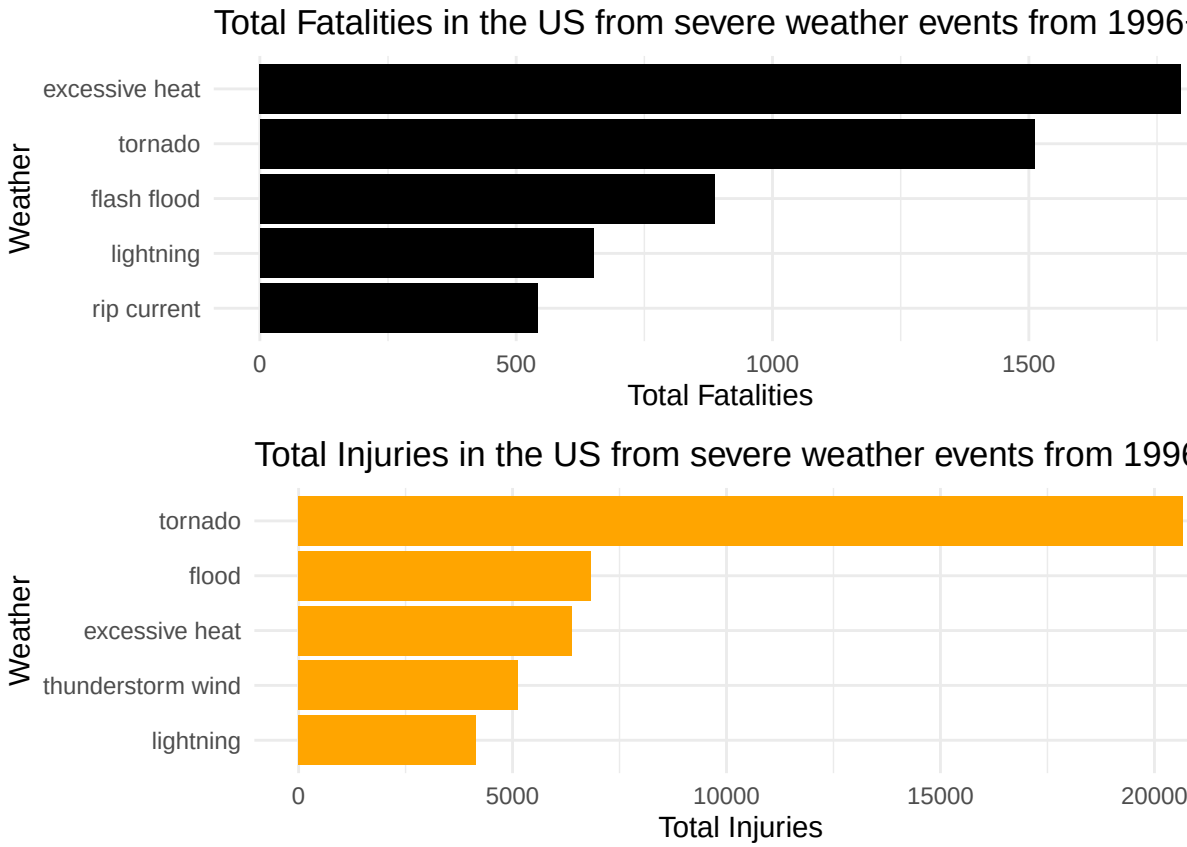
Q1) Across the United States, which types of events (as indicated in the EVTYPE variable) are most harmful with respect to population health?

```
storm.data.fatality<-ddply(storm.data.event,
                           .(evtype),
                           summarise,
                           total_fatalities=sum(fatalities))

storm.data.fatality<-storm.data.fatality[order(-storm.data.fatality$total_fatalities),]

storm.data.fatality.10<-head(storm.data.fatality,10)

storm.data.injury<-ddply(storm.data.event,
                         .(evtype),
                         summarise,
                         total_injuries=sum(injuries))

storm.data.injury<-storm.data.injury[order(-storm.data.injury$total_injuries),]

storm.data.injury.10<-head(storm.data.injury,10)

storm.data.fat_inj<-ddply(storm.data.event,
                         .(evtype),
                         summarise,
                         total_harm=sum(injuries)+sum(fatalities))

storm.data.fat_inj<-storm.data.fat_inj[order(-storm.data.fat_inj$total_harm),]

storm.data.fat_inj.10<-head(storm.data.fat_inj,10)
```

setting up the datasets by grouping them by evtype and ordering them in *descending order*

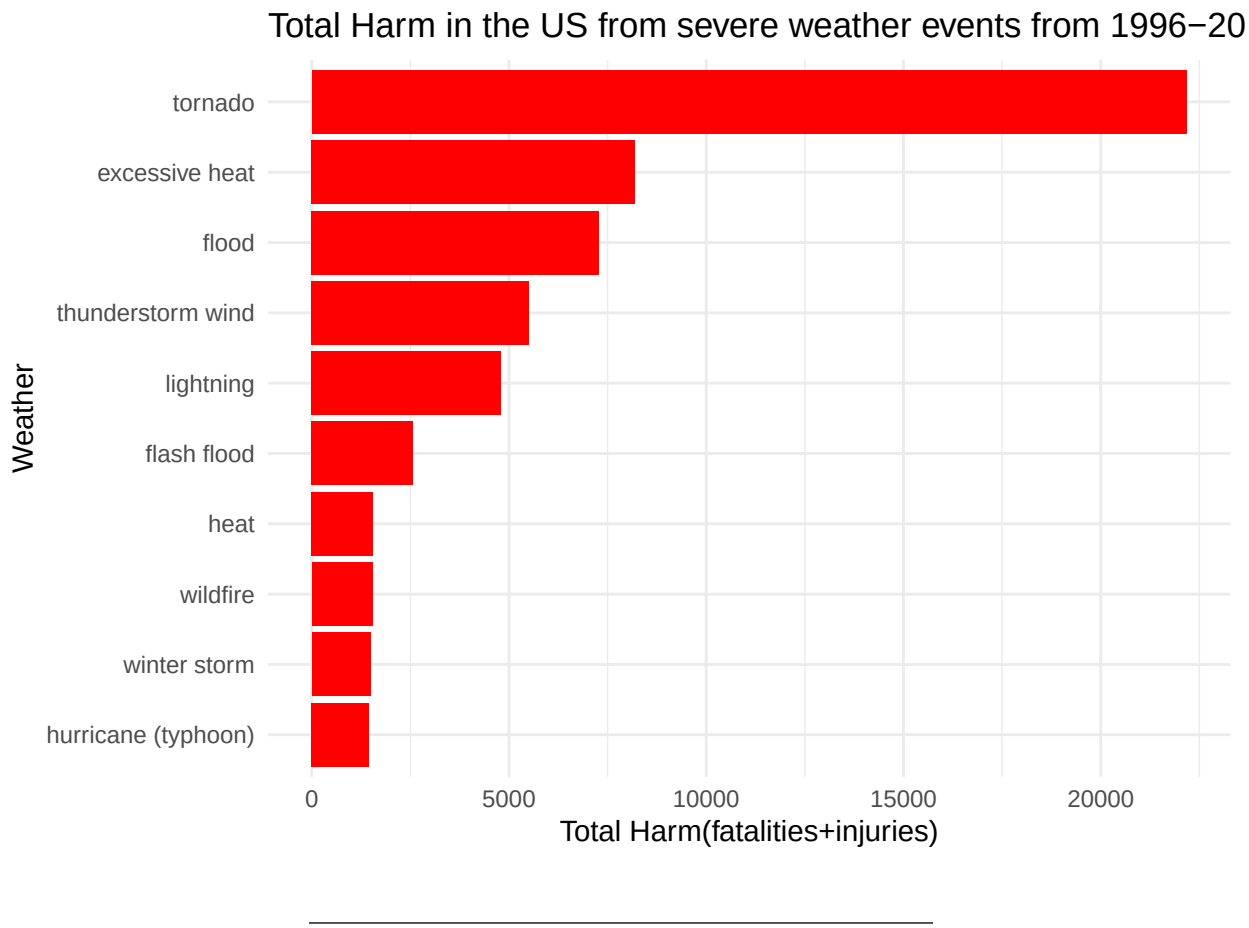
```
plot_fatalities<-ggplot(storm.data.fatality.10[1:5,], aes(reorder(evtype, total_fatalities), total_fatal
  geom_col(fill = "black") + coord_flip() + labs(y = "Total Fatalities",
       x = "Weather",
       title = "Total Fatalities in the US from severe weather events from 1996-2011")+ theme_minimal()
```

```
plot_injuries<-ggplot(storm.data.injury.10[1:5,], aes(reorder(evtype, total_injuries), total_injuries))
  geom_col(fill = "orange") + coord_flip() + labs(y = "Total Injuries",
       x = "Weather",
       title = "Total Injuries in the US from severe weather events from 1996-2011")+ theme_minimal()

plot_all<-ggplot(storm.data.fat_inj.10, aes(reorder(evtype, total_harm), total_harm)) +
  geom_col(fill = "red") + coord_flip() + labs(y = "Total Harm(fatalities+injuries)",
       x = "Weather",
       title = "Total Harm in the US from severe weather events from 1996-2011")+ theme_minimal()

plot_grid(plot_fatalities, plot_injuries, plot_all,
          ncol = 1, nrow = 2)
```



plotting the data

```
plot_all
```



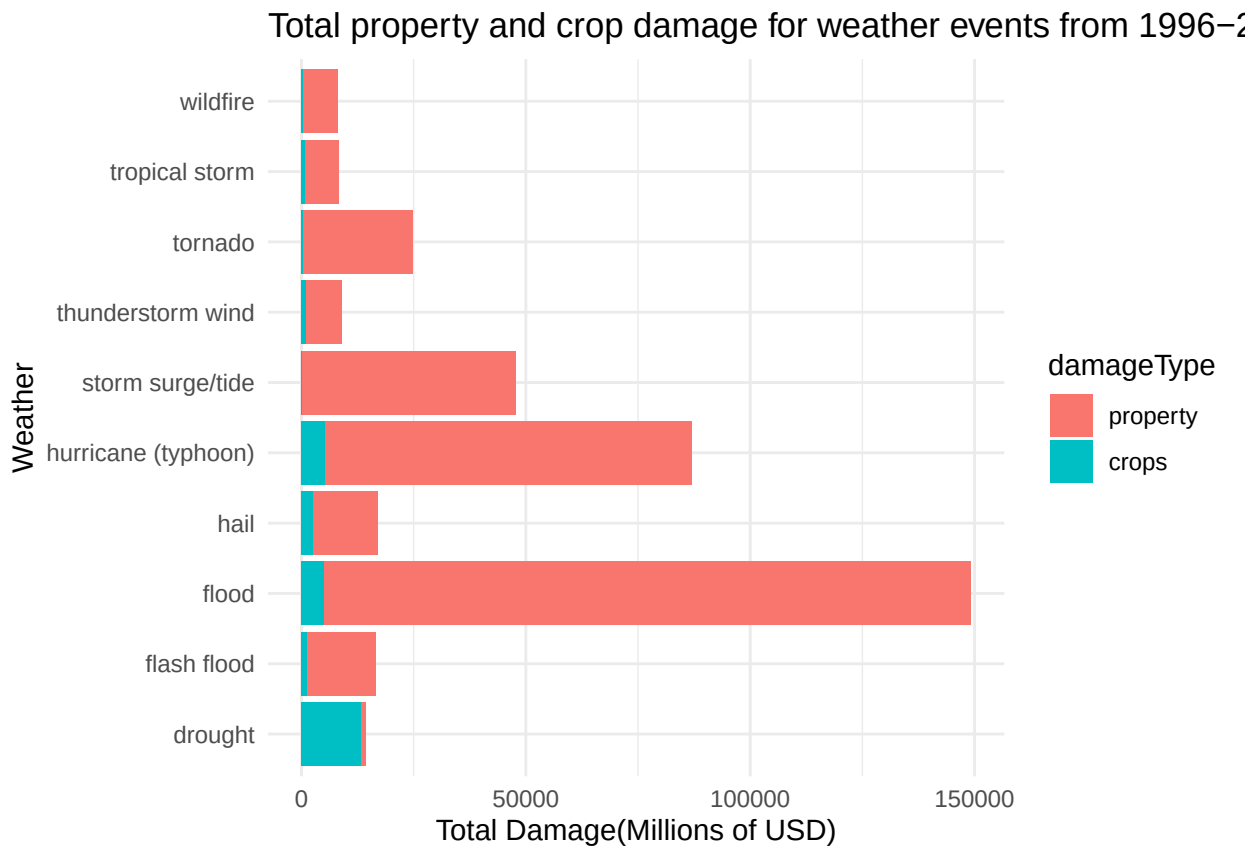
Q2) Across the United States, which types of events have the greatest economic consequences?

```
results.economic<-ddply(storm.data.event,
                        .(evtype),
                        summarise,
                        propdmg_total = sum(propdmg_total),
                        cropdmg_total = sum(cropdmg_total),
                        dmg_total=propdmg_total+cropdmg_total)

results.economic<-results.economic[order(-results.economic$dmg_total),]

results.economic.10<-results.economic[1:10,]

results.economic.10melt<- melt(results.economic.10,
                               id.vars=c("evtype"),
                               measure.vars=c("propdmg_total","cropdmg_total"),
                               variable.name="damageType",
                               value.name="damage")

levels(results.economic.10melt$damageType)[levels(results.economic.10melt$damageType)=="propdmg_total"]
levels(results.economic.10melt$damageType)[levels(results.economic.10melt$damageType)=="cropdmg_total"]
```

melt the dataset and categorise the damage by *propdmg* for property damage and *cropdmg* for the cropdamage

```
ggplot(results.economic.10melt, aes(x=evtype, y=damage/1000000)) + geom_bar(stat = "identity", aes(fill=
```

Plot the result



Results

From the above analysis we answer the following questions:-

Q1) Across the United States, which types of events (as indicated in the **EVTYPE** variable) are most harmful with respect to population health?

Ans 1) Across the United States, from 1996-2011, *excessive heat* causes most *fatalities tornado* causes most *injuries* **tornado** causes most harm population health on the aggregrate(injuries+fatalities)

Q2) Across the United States, which types of events have the greatest economic consequences?

Ans 2) Across the United States, from 1996-2011, the top 4 events which have the greatest economic consequences are **flood** , **hurricane(typhoon)** , **storm surge/tide** , **tornado**