# Health and Economic Outcomes of Weather Events 1951-2011: An Overview

ISY

12/21/2020

## Synopsis

Storms and other severe weather events can cause both public health and economic problems for communities and municipalities. In this research, data from US National Oceanic and Atmospheric Administration had been obtained and processed to see the outcomes by each event. Data were obtained (link in loading and processing the raw data section), cleaned (data processing section) and summarized (pre-analysis section). Results have been shown in a respective section. Data were analysed in RStudio R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

## Loading and Processing the Raw Data

From the EPA Air Quality System, we obtained data particulate matter air pollution levels that is monitored accros the USA by a nationwide PM monitoring source. We obtained files that presents the data from 1999 and 2012.

### Data Processing

Dataset were given in this link in the assignment instructions.

```
#Set directories and download dataset

directory <- getwd()
urldata <- "https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2"
filepath <- paste0(directory,"/stormdata.csv.bz2")

if (!file.exists(filepath)) {download.file(urldata, destfile = filepath)}

#Download Data Documentation
urldoc <- "https://d396qusza40orc.cloudfront.net/repdata%2Fpeer2_doc%2Fpd01016005curr.pdf"
docpath <- paste0(directory,"/datadoc.pdf")
if (!file.exists(docpath)) {download.file(urldoc, destfile = docpath)}

storm <- read.csv(file = filepath, na.strings = NA)

#Summarize the dataset

str(storm)
```

```
## 'data.frame':    902297 obs. of  37 variables:
##  $ STATE__   : num  1 1 1 1 1 1 1 1 1 1 ...
##  $ BGN_DATE  : chr  "4/18/1950 0:00:00" "4/18/1950 0:00:00" "2/20/1951 0:00:00" "6/8/1951 0:00:00" .
```

```
##  $ BGN_TIME  : chr  "0130" "0145" "1600" "0900" ...
##  $ TIME_ZONE : chr  "CST" "CST" "CST" "CST" ...
##  $ COUNTY    : num  97 3 57 89 43 77 9 123 125 57 ...
##  $ COUNTYNAME: chr  "MOBILE" "BALDWIN" "FAYETTE" "MADISON" ...
##  $ STATE     : chr  "AL" "AL" "AL" "AL" ...
##  $ EVTYPE    : chr  "TORNADO" "TORNADO" "TORNADO" "TORNADO" ...
##  $ BGN_RANGE : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ BGN_AZI   : chr  "" "" "" "" ...
##  $ BGN_LOCATI: chr  "" "" "" "" ...
##  $ END_DATE  : chr  "" "" "" "" ...
##  $ END_TIME  : chr  "" "" "" "" ...
##  $ 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   : chr  "" "" "" "" ...
##  $ END_LOCATI: chr  "" "" "" "" ...
##  $ 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: chr  "K" "K" "K" "K" ...
##  $ CROPDMG   : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ CROPDMGEXP: chr  "" "" "" "" ...
##  $ WFO       : chr  "" "" "" "" ...
##  $ STATEOFFIC: chr  "" "" "" "" ...
##  $ ZONENAMES : chr  "" "" "" "" ...
##  $ 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   : chr  "" "" "" "" ...
##  $ REFNUM    : num  1 2 3 4 5 6 7 8 9 10 ...
```

**Pre-analizing Processing**

These are the questions that are asked in the assignment:

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?

Since our dataset has a relatively big size, required information will be selected from the dataset and a new dataset will be created. The information should be on health and economic assets.

As we check from the data documentation, the following columns in *storm* dataset contains information to regarding:

*EVTYPE: Event Type* FATALITIES : Mortality *INJURIES : # of Injuries* PRODMG : Property Damage in US Dollars *PROPDMGEXP: The units property damage* CROPDMG : Crop Damage in US Dollars *CROPDMGEXP: The units for property damage

*dplyr*, *data.table* and *lubridate* packages will be used for this pre-analizing process.

```
library(dplyr)
library(data.table)
library(lubridate)

vars <- c("EVTYPE", "FATALITIES", "INJURIES", "PROPDMG", "PROPDMGEXP","CROPDMG","CROPDMGEXP")
mydata <- storm[, vars]
```

Check for information on missing values.

```
for (i in colnames(mydata)) {print(sum(is.na(mydata[[i]])))}
```

```
## [1] 0
## [1] 0
## [1] 0
## [1] 0
## [1] 0
## [1] 0
## [1] 0
```

Event types might vary, We wil check for any combination possibilities.

```
sort(table(mydata$EVTYPE), decreasing = TRUE)[1:20]
```

```
##
##                      HAIL                 TSTM WIND         THUNDERSTORM WIND
##                    288661                    219940                     82563
##                   TORNADO               FLASH FLOOD                    FLOOD
##                     60652                     54277                     25326
##       THUNDERSTORM WINDS                 HIGH WIND                 LIGHTNING
##                     20843                     20212                     15754
##                 HEAVY SNOW                HEAVY RAIN               WINTER STORM
##                     15708                     11723                     11433
##            WINTER WEATHER               FUNNEL CLOUD          MARINE TSTM WIND
##                      7026                      6839                      6175
## MARINE THUNDERSTORM WIND                 WATERSPOUT               STRONG WIND
##                      5812                      3796                      3566
##     URBAN/SML STREAM FLD                  WILDFIRE
##                      3392                      2761
```

We will rename the event types to their main categories in terms of presentation purposes. If events are relatively small in numbers, we will assign them as "Other". Regrouping will be done by extracting values to a new variable: "EVENT". Note that we will focus on storm related info.

```
mydata$EVENT <- "Other"

mydata$EVENT[grep("HAIL", mydata$EVTYPE, ignore.case = TRUE)] <- "Hail"
mydata$EVENT[grep("HEAT", mydata$EVTYPE, ignore.case= TRUE)] <- "Heat"
mydata$EVENT[grep("FLOOD", mydata$EVTYPE, ignore.case = TRUE)] <- "Flood"
mydata$EVENT[grep("WIND", mydata$EVTYPE, ignore.case = TRUE)] <- "Wind"
mydata$EVENT[grep("STORM", mydata$EVTYPE, ignore.case= TRUE)] <- "Storm"
mydata$EVENT[grep("SNOW", mydata$EVTYPE, ignore.case = TRUE)] <- "Snow"
mydata$EVENT[grep("TORNADO", mydata$EVTYPE, ignore.case = TRUE)] <- "Tornado"
mydata$EVENT[grep("WINTER",mydata$EVTYPE, ignore.case= TRUE)] <- "Winter"
mydata$EVENT[grep("RAIN", mydata$EVTYPE, ignore.case = TRUE)] <- "Rain"

#Check for newly created variable.
```

```
sort(table(mydata$EVENT), decreasing = TRUE)
```

```
##
##    Hail    Wind   Storm   Flood Tornado   Other  Winter    Snow    Rain    Heat
##  289270  255362  113156   82686   60700   48970   19604   17660   12241    2648
```

Check for distinct values for expedition data.

```
for (i in c("PROPDMGEXP","CROPDMGEXP")) {print(unique(mydata[[i]]))}
```

```
##  [1] "K" "M" ""  "B" "m" "+" "0" "5" "6" "?" "4" "2" "3" "h" "7" "H" "-" "1" "8"
## [1] ""  "M" "K" "m" "B" "?" "0" "k" "2"
```

Some of these values mean:

*K: Thousand dollars (10^3)* B: Billion dollars (10^9) *M: Million dollars (10^6)

Rest will be considered as dollars.

```
#Check for damage variables class. If anything different than character, convert to character

mydata$PROPDMGEXP <- as.character(mydata$PROPDMGEXP)
mydata$CROPDMGEXP <- as.character(mydata$CROPDMGEXP)

#Change property expedition values to numeric, multiply the new values with
#property damage value and store the new values in a new variable.

mydata$PROPDMGEXP[grep("K", mydata$PROPDMGEXP, ignore.case = TRUE)] <- 3
mydata$PROPDMGEXP[grep("M", mydata$PROPDMGEXP, ignore.case = TRUE)] <- 6
mydata$PROPDMGEXP[grep("B", mydata$PROPDMGEXP, ignore.case = TRUE)] <- 9
mydata$PROPDMGEXP[grep("K|M|B", mydata$PROPDMGEXP, ignore.case= TRUE)] <- 0

mydata$PROPDMGEXP <- as.numeric(mydata$PROPDMGEXP)
mydata$PROPDMGEXP[is.na(mydata$PROPDMGEXP)] <- 0
mydata$PROP.DMG <- mydata$PROPDMG * 10^mydata$PROPDMGEXP
```

```
# Change crop damage expeditions to numeric, multiply the new values with crop damage
#values and store the new values in a new variable

mydata$CROPDMGEXP[grep("K", mydata$CROPDMGEXP, ignore.case = TRUE)] <- 3
mydata$CROPDMGEXP[grep("M", mydata$CROPDMGEXP, ignore.case = TRUE)] <- 6
mydata$CROPDMGEXP[grep("B", mydata$CROPDMGEXP, ignore.case = TRUE)] <- 9
mydata$CROPDMGEXP[grep("K|M|B", mydata$CROPDMGEXP, ignore.case = TRUE)] <- 0

mydata$CROPDMGEXP <- as.numeric(mydata$CROPDMGEXP)
mydata$CROPDMGEXP[is.na(mydata$CROPDMGEXP)] <- 0
mydata$CROP.DMG <- mydata$CROPDMGEXP * 10^mydata$CROPDMGEXP

#Check the new variables

str(mydata$PROP.DMG)
```

```
##  num [1:902297] 25000 2500 25000 2500 2500 2500 2500 2500 25000 25000 ...
```

```
str(mydata$CROP.DMG)
```

```
##  num [1:902297] 0 0 0 0 0 0 0 0 0 0 ...
```

Check for any NA's in the dataset

```
for (i in colnames(mydata)) {print(sum(is.na(mydata[[i]])))}
```

```
## [1] 0
## [1] 0
## [1] 0
## [1] 0
## [1] 0
## [1] 0
## [1] 0
## [1] 0
## [1] 0
## [1] 0
```

Since we need the the outcomes for both health and economics, we will create a data frame that contains the summary of both health and economics

```
outcomes <- mydata %>% group_by(EVENT) %>%
     summarize(Fatalities = sum(FATALITIES),
               Injuries = sum(INJURIES),
               totalHealth = sum(sum(FATALITIES), sum(INJURIES)),
               PropDMG = sum(PROP.DMG),
               CropDMG= sum(CROP.DMG),
               totalEcon = sum(sum(PROP.DMG),sum(CROP.DMG)))

# Add total cases
totalCases <- as.data.frame(table(mydata$EVENT))[,2]
outcomes <- cbind(outcomes, totalCases)

#Order the columns for better exploration
outcomes <- outcomes[, c(1,8,2:7)]
```

## Analysis

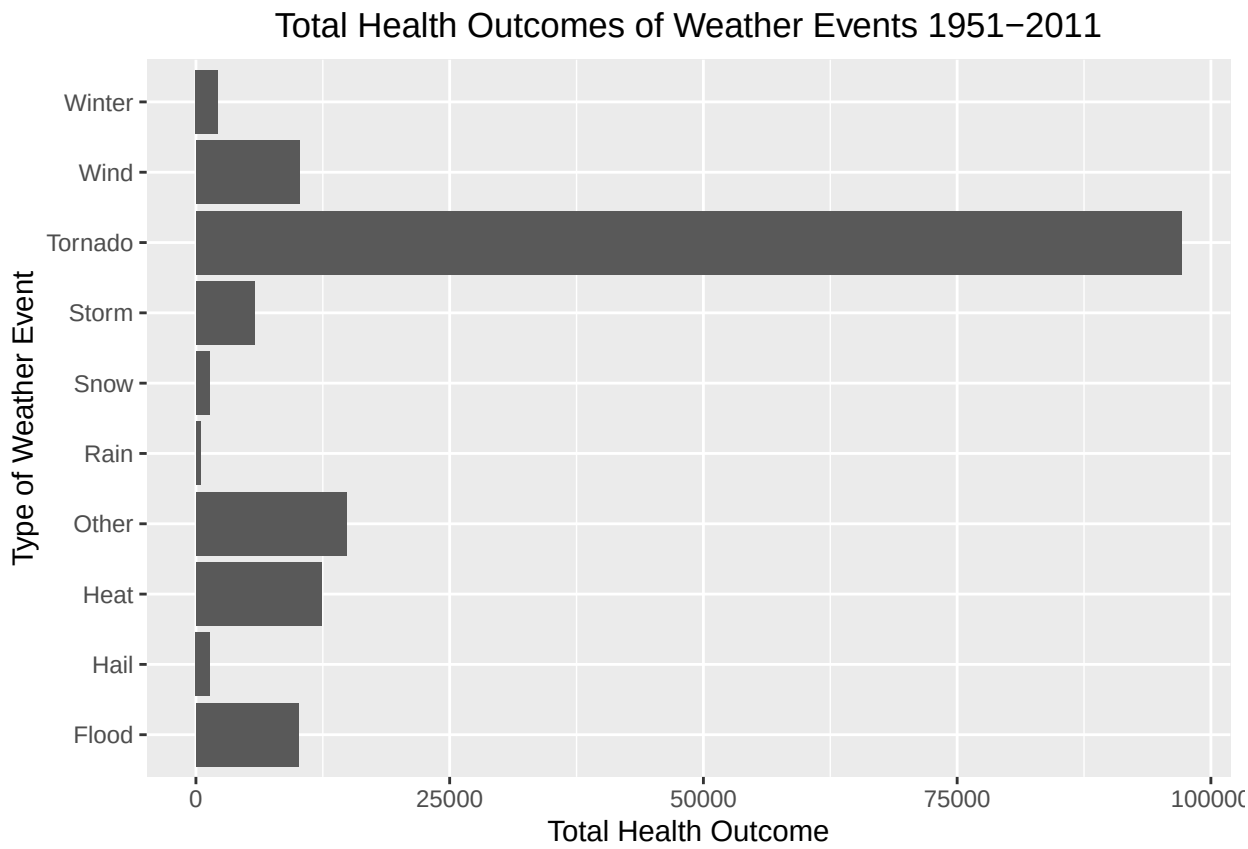
### Part 1: Public Health Outcomes

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

We have created an *outcomes* table before. We will create a barplot to show the health outcomes. *ggplot2* package will be used.

```
library(ggplot2)
plot1  <- ggplot(data = outcomes, aes(x=totalHealth, y=EVENT), ) +
     geom_bar(stat="identity") +
     xlab("Total Health Outcome") +
     ylab("Type of Weather Event") +
     ggtitle("Total Health Outcomes of Weather Events 1951-2011") +
     theme(plot.title = element_text(hjust = 0.5))

plot1
```




As a result, worst health outcomes had been seen because of tornados.

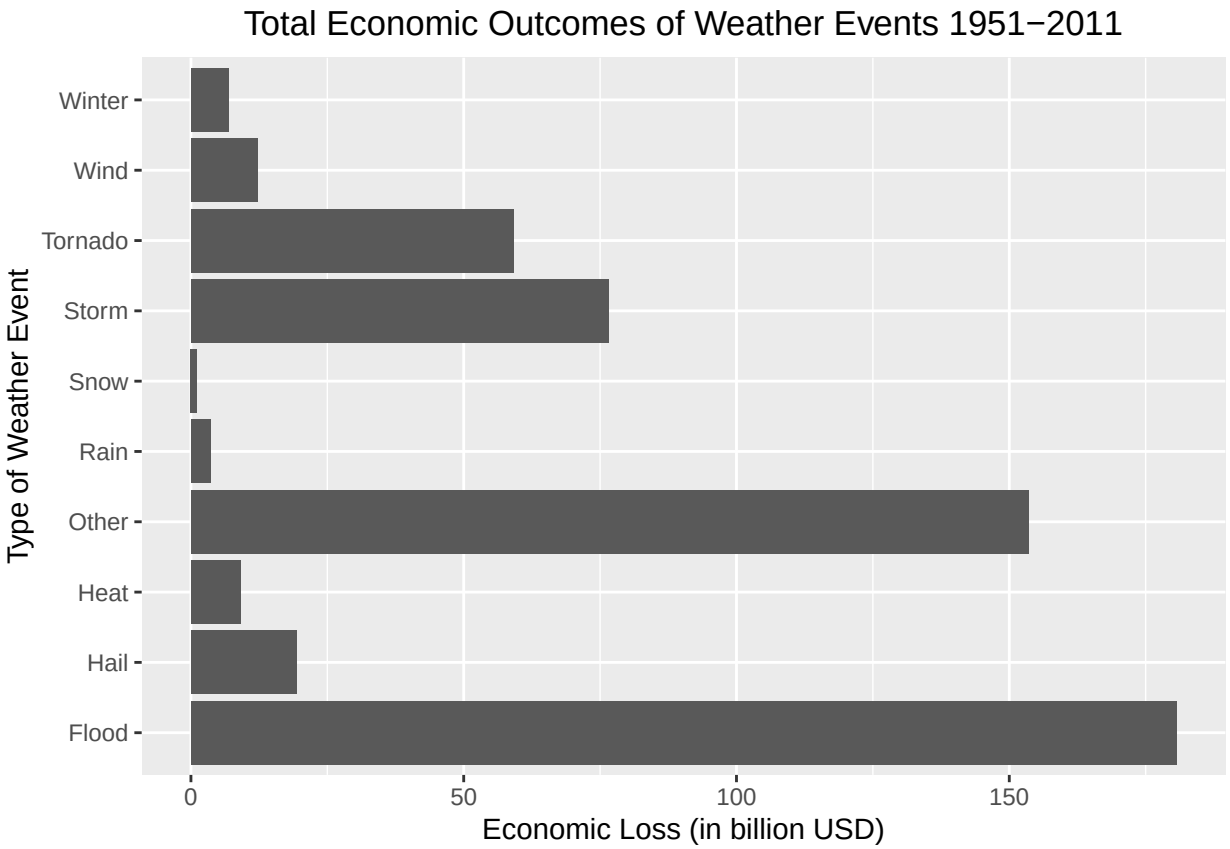
**Part 2: Economic outcomes**

Question is: Across the United States, which types of events have the greatest economic consequences?

*outcomes* dataset will be used.

```
plot2 <- ggplot(data= outcomes, aes(x= totalEcon/1E9, y= EVENT)) +
     geom_histogram(stat = "identity") +
     xlab("Economic Loss (in billion USD)") +
     ylab("Type of Weather Event") +
     ggtitle("Total Economic Outcomes of Weather Events 1951-2011") +
     theme(plot.title = element_text(hjust = 0.5))

plot2
```




As a result, we can say that most devastating event for property and corps is flood. (excluding "others")