NOAA Storm Database Exploration for Health Impact and Economic Cost

AK

09/25/2015

Synopsis

The objective of this report is to present the analysis done on the NOAA Storm database to determine the health impact and economic costs of severe weather events such as tornodoes, thundurstorms, earthquake etc and determine the weather event that has the most impact. The harm to the population health was calculated as a sum of the fatalities and injuries while the economic cost was calculated as a sum of the property damage and crop damage in dollars (after factoring in the exponents post some clean-up). The entire analysis (including data download, pre-processing, clean-up and exploratory analysis) was done programatically for complete reproducibility using publicly available R-packages. The complete historically data (total records: 902297) was considered in the analysis without any specific weightings for the time and place of the data capture. Though there are a total of 985 weather events captured in the database since 1950, only 48 weather events are currently prevalent. Some basic clean-up relevant to address the questions was done to normalize these events as most others are non-consequential. Only top 25 events sorted by health impact and economic cost have been summarized in the figures for ease of presentation and review.

Results:

Based on the analysis, it was determined that:

- **Tornadoes are most harmful to the health of the population. Tornadoes cause a total of 5658 fatalities and 91364 injuries (total health impact: 97022).**
  Refer to Plot 1 and Table 1 to review the health impact of top 25 events.
- **Tornadoes are most impactful in terms of the economic cost. Tornadoes cause a total of 43,631,180,105 (~42Bn) in total cost (property damage: 43,487,203,105 and crop damage: 143,977,000**
  Refer to Plot 2 and Table 2 to review the economic cost of top 25 events.
- **The findings for the top 5 w.r.t health impact and the economic cost seem logical (related to tornadoes, storms, floods, heat and winds)**

Data Processing

```
        ##Setting the global options
        suppressWarnings(library(knitr))
        opts_chunk$set(echo = TRUE)
        opts_chunk$set(fig.path = "./figures/")
```

```
        ##Load all the required libraries for the program
        ##Ignore the warnings as those have been checked to be harmless
        suppressWarnings(library(lubridate)) ##Required for date manipulation
        suppressWarnings(library(reshape2)) ##For melt and cast
        suppressWarnings(library(lattice)) ##Will use Lattice Graphing system for the las
t plot
        suppressWarnings(suppressMessages(library(dplyr)))##Required to mutate
        suppressWarnings(library(xtable)) ##To output tables
```

```
        ##---Start of Code for downloading, reading and preprocessing data ---

        ##Download and unzip the data file
        zippeddataURL<-("https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormDat
a.csv.bz2")

        ##Download the Zip File only if it is not already there.
        zippeddatafile <-"StormData.csv.bz2"
        if (!file.exists(zippeddatafile)){
                download.file(zippeddataURL, zippeddatafile)
                dateDownloaded<-date()
        }

        ##Now read the file in with the right parameters.
        stormdata <-read.table(zippeddatafile, header=TRUE, sep=",", na.strings="NA",
                               stringsAsFactors=FALSE)

        ##Total number of entries
        entries<-nrow(stormdata) ##902297. So file read correctly.

        ##Let us first see how many unique event types are in the storm data
        eventtypesindata<-unique(stormdata$EVTYPE)

        ##---End of Code for downloading, reading and preprocessing data ---
```

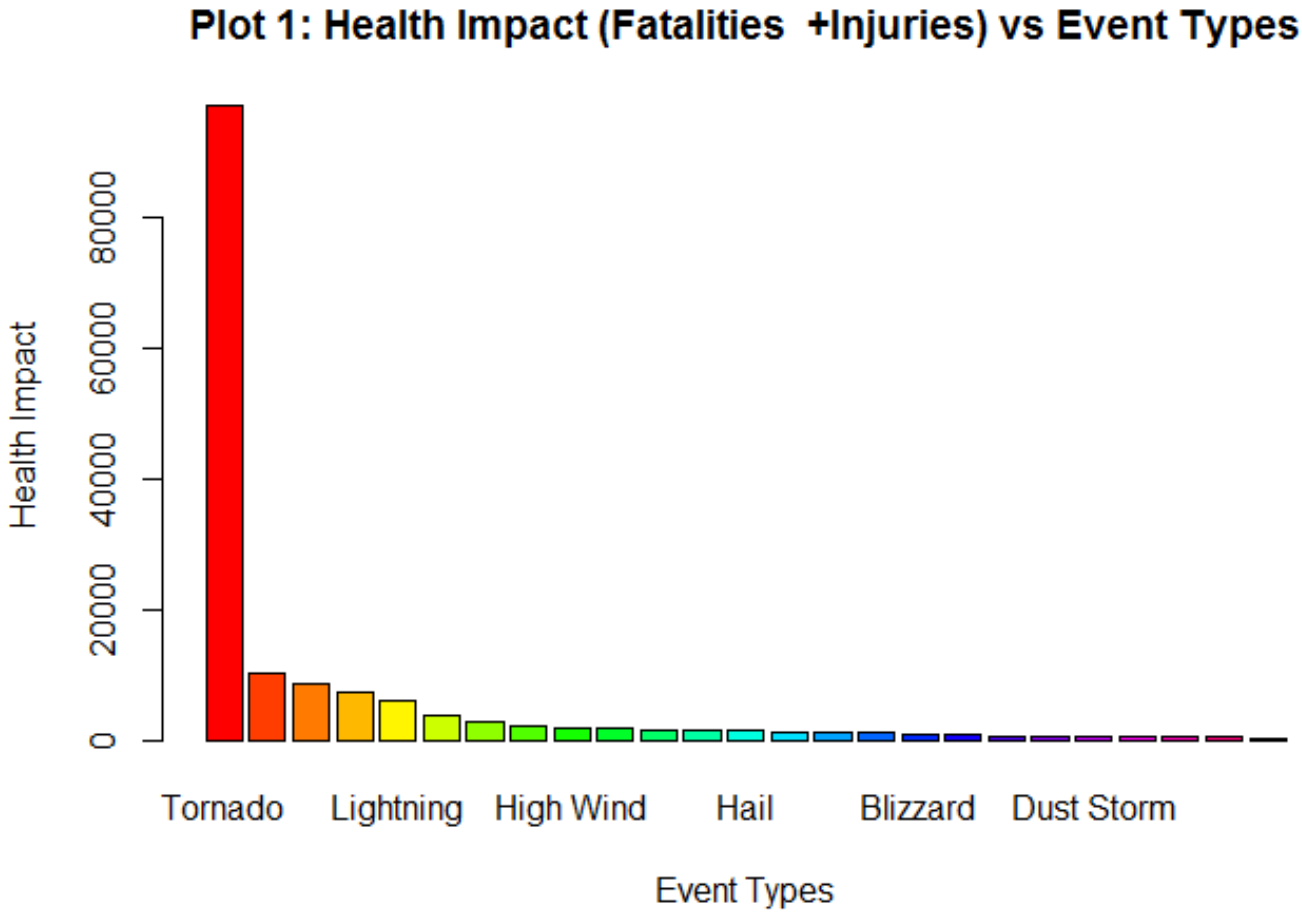
- The total number of records in the NOAA storm data is **902297**
- The total number of initial unique event types in the storm data is **985**

```
        ##Normalize the PROPEXP and CROPEXP columns.
  uniqueexp<-unique(stormdata$PROPDMGEXP)
        ##[1] "K" "M" ""  "B" "m" "+" "0" "5" "6" "?" "4" "2" "3" "h" "7" "H" "-" "1" "8"

        uniqueexp<-unique(stormdata$CROPDMGEXP)
        ##[1] ""  "M" "K" "m" "B" "?" "0" "k" "2"

        ##It is clear that there are some additional values. Data operators seem to have
interpreted exponents as generic rather than just use K, M, B. So 2 is x100, 3 is x1000,
4 is x10000 etc.
        stormdata$PROPDMGEXP[which(stormdata$PROPDMGEXP=="H")]<-2
        stormdata$PROPDMGEXP[which(stormdata$PROPDMGEXP=="h")]<-2

        stormdata$PROPDMGEXP[which(stormdata$PROPDMGEXP=="K")]<-3
        stormdata$PROPDMGEXP[which(stormdata$PROPDMGEXP=="k")]<-3

        stormdata$PROPDMGEXP[which(stormdata$PROPDMGEXP=="M")]<-6
        stormdata$PROPDMGEXP[which(stormdata$PROPDMGEXP=="m")]<-6

        stormdata$PROPDMGEXP[which(stormdata$PROPDMGEXP=="B")]<-9
        stormdata$PROPDMGEXP[which(stormdata$PROPDMGEXP=="b")]<-9

        stormdata$PROPDMGEXP[which(stormdata$PROPDMGEXP=="")]<-0
        stormdata$PROPDMGEXP[which(stormdata$PROPDMGEXP=="?")]<-0
        stormdata$PROPDMGEXP[which(stormdata$PROPDMGEXP=="+")]<-0
        stormdata$PROPDMGEXP[which(stormdata$PROPDMGEXP=="-")]<-0

        stormdata$CROPDMGEXP[which(stormdata$CROPDMGEXP=="H")]<-2
        stormdata$CROPDMGEXP[which(stormdata$CROPDMGEXP=="h")]<-2

        stormdata$CROPDMGEXP[which(stormdata$CROPDMGEXP=="K")]<-3
        stormdata$CROPDMGEXP[which(stormdata$CROPDMGEXP=="k")]<-3

        stormdata$CROPDMGEXP[which(stormdata$CROPDMGEXP=="M")]<-6
        stormdata$CROPDMGEXP[which(stormdata$CROPDMGEXP=="m")]<-6

        stormdata$CROPDMGEXP[which(stormdata$CROPDMGEXP=="B")]<-9
        stormdata$CROPDMGEXP[which(stormdata$CROPDMGEXP=="b")]<-9


        stormdata$CROPDMGEXP[which(stormdata$CROPDMGEXP=="")]<-0
        stormdata$CROPDMGEXP[which(stormdata$CROPDMGEXP=="?")]<-0
        stormdata$CROPDMGEXP[which(stormdata$CROPDMGEXP=="+")]<-0
        stormdata$CROPDMGEXP[which(stormdata$CROPDMGEXP=="-")]<-0

        ##This standard event type list is retrieved from the Storm Data Event Table on P
age 6
        ##https://d396qusza40orc.cloudfront.net/repdata%2Fpeer2_doc%2Fpd01016005curr.pdf

        stdeventypes<-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 Sur
f",
        "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 Stor
m",
        "Tsunami","Volcanic Ash","Waterspout","Wildfire","Winter Storm","Winter Weather")

        ##Pass 1: Correct event types that start with the standard ones above
        for (i in 1:length(stdeventypes)){
                stormdata$EVTYPE[grep(paste0("^", stdeventypes[i]), stormdata$EVTYPE, ign
ore.case=T)]<-stdeventypes[i]
        }

        ##Pass 2: Correct the few top events
        stormdata$EVTYPE[grep("^TSTM WIND",stormdata$EVTYPE, ignore.case=T)]<- "Thunderst
orm Wind"
        stormdata$EVTYPE[grep("^FOG",stormdata$EVTYPE, ignore.case=T)]<- "Dense Fog"
        stormdata$EVTYPE[which(stormdata$EVTYPE=="WILD/FOREST FIRE")]<-"Wildfire"
        stormdata$EVTYPE[which(stormdata$EVTYPE=="WILD FIRES")]<-"Wildfire"
        stormdata$EVTYPE[which(stormdata$EVTYPE=="HEAT WAVE")]<-"Excessive Heat"
        stormdata$EVTYPE[which(stormdata$EVTYPE=="EXTREME HEAT")]<-"Excessive Heat"
        stormdata$EVTYPE[which(stormdata$EVTYPE=="STORM SURGE")]<-"Storm Surge/Tide"
        stormdata$EVTYPE[which(stormdata$EVTYPE=="TYPHOON")]<-"Hurricane/Typhoon"
        stormdata$EVTYPE[which(stormdata$EVTYPE=="WINTRY MIX")]<-"Winter Weather"
        stormdata$EVTYPE[which(stormdata$EVTYPE=="GLAZE")]<-"Sleet"
        stormdata$EVTYPE[grep("^HURRICANE",stormdata$EVTYPE, ignore.case=T)]<- "Hurrican
e/Typhoon"
        stormdata$EVTYPE[grep("^GUSTY WIND",stormdata$EVTYPE, ignore.case=T)]<- "Strong W
ind"
        stormdata$EVTYPE[grep("^WIND",stormdata$EVTYPE, ignore.case=T)]<- "High Wind"
        stormdata$EVTYPE[which(stormdata$EVTYPE=="EXTREME COLD")]<-"Extreme Cold/Wind Chi
ll"
        stormdata$EVTYPE[which(stormdata$EVTYPE=="EXCESSIVE RAINFALL")]<-"Heavy Rain"
        stormdata$EVTYPE[grep("^FREEZE",stormdata$EVTYPE, ignore.case=T)]<- "Frost/Freez
e"
        stormdata$EVTYPE[grep("^SNOW",stormdata$EVTYPE, ignore.case=T)]<- "Heavy Snow"
        stormdata$EVTYPE[grep("^ICE",stormdata$EVTYPE, ignore.case=T)]<- "Ice Storm"

        ##Assuming that Landslies are caused by Heavy Rains
        stormdata$EVTYPE[grep("^LANDSLIDE",stormdata$EVTYPE, ignore.case=T)]<-"Heavy Rai
n"
```

Q1: Across the United States, which types of events are most harmful with respect to population health?

```
        ##---Start of Code to Answer Q1 ---
        ##Using storm data frame from loaddata r code chunk above. Remove the NAs. Retain
values that are non-zero.  This is being done given the question to be answered w.r.t fat
alities and injuries (harmful to population health)
        stormdata<-subset(stormdata, !is.na(FATALITIES) & !is.na(INJURIES) & (FATALITIE
S>0 | INJURIES>0))

        ##One can use group by approach below but I like melt and cast
        ##Now melt the data set with event type as ID.
        meltedstormdatabyevent <- melt(stormdata, id.vars = c("EVTYPE"), measure.var
s=c("FATALITIES", "INJURIES"))

        ##Then cast the data set with event type as ID and calculate the average by event
type
        castedstormdatabyevent <- dcast(meltedstormdatabyevent, EVTYPE~variable,sum)

        ##Introduce a column HEALTHIMPACT that is a sum of FATALITIES and INJURIES
        castedstormdatabyevent <- mutate(castedstormdatabyevent, HEALTHIMPACT=round(FATAL
ITIES+INJURIES))

        ##Order by HEALTHIMPACT
        castedstormdatabyevent<-arrange(castedstormdatabyevent,desc(HEALTHIMPACT))

        ##Get the max health impact and the corresponding event
        maxheathimpact<-max(castedstormdatabyevent$HEALTHIMPACT)
        eventtypewithmaxheathimpact<-castedstormdatabyevent[which(castedstormdatabyeven
t$HEALTHIMPACT==maxheathimpact), ][1,1]

        ##Plot and print the table for Top 25 events only
        top25<-head(castedstormdatabyevent, 25)
        par(mfcol=c(1,1))
        barplot(top25$HEALTHIMPACT, names.arg=top25$EVTYPE, main="Plot 1: Health Impact
(Fatalities  +Injuries) vs Event Types",
          xlab="Event Types", ylab="Health Impact", col=rainbow(25))
```




```
        healthimpact<-xtable(top25, caption="Table 1: Health Impact for top 25 event type
s")
        print(healthimpact, type="html")
```

	EVTYPE	FATALITIES	INJURIES	HEALTHIMPACT
1	Tornado	5658.00	91364.00	97022.00
2	Thunderstorm Wind	709.00	9458.00	10167.00
3	Excessive Heat	1999.00	6680.00	8679.00
4	Flood	495.00	6806.00	7301.00
5	Lightning	817.00	5232.00	6049.00
6	Heat	1118.00	2494.00	3612.00
7	Flash Flood	1018.00	1785.00	2803.00
8	Ice Storm	96.00	2115.00	2211.00
9	High Wind	318.00	1558.00	1876.00
10	Wildfire	90.00	1606.00	1696.00
11	Winter Storm	217.00	1353.00	1570.00
12	Hurricane/Typhoon	135.00	1333.00	1468.00
13	Hail	15.00	1361.00	1376.00
14	Heavy Snow	142.00	1137.00	1279.00
15	Dense Fog	81.00	1077.00	1158.00

16	Rip Current	577.00	529.00	1106.00
17	Blizzard	101.00	805.00	906.00
18	Winter Weather	62.00	615.00	677.00
19	Extreme Cold/Wind Chill	285.00	255.00	540.00
20	Heavy Rain	139.00	329.00	468.00
21	Dust Storm	22.00	440.00	462.00
22	Tropical Storm	66.00	383.00	449.00
23	Strong Wind	116.00	313.00	429.00
24	Avalanche	224.00	170.00	394.00
25	High Surf	104.00	156.00	260.00

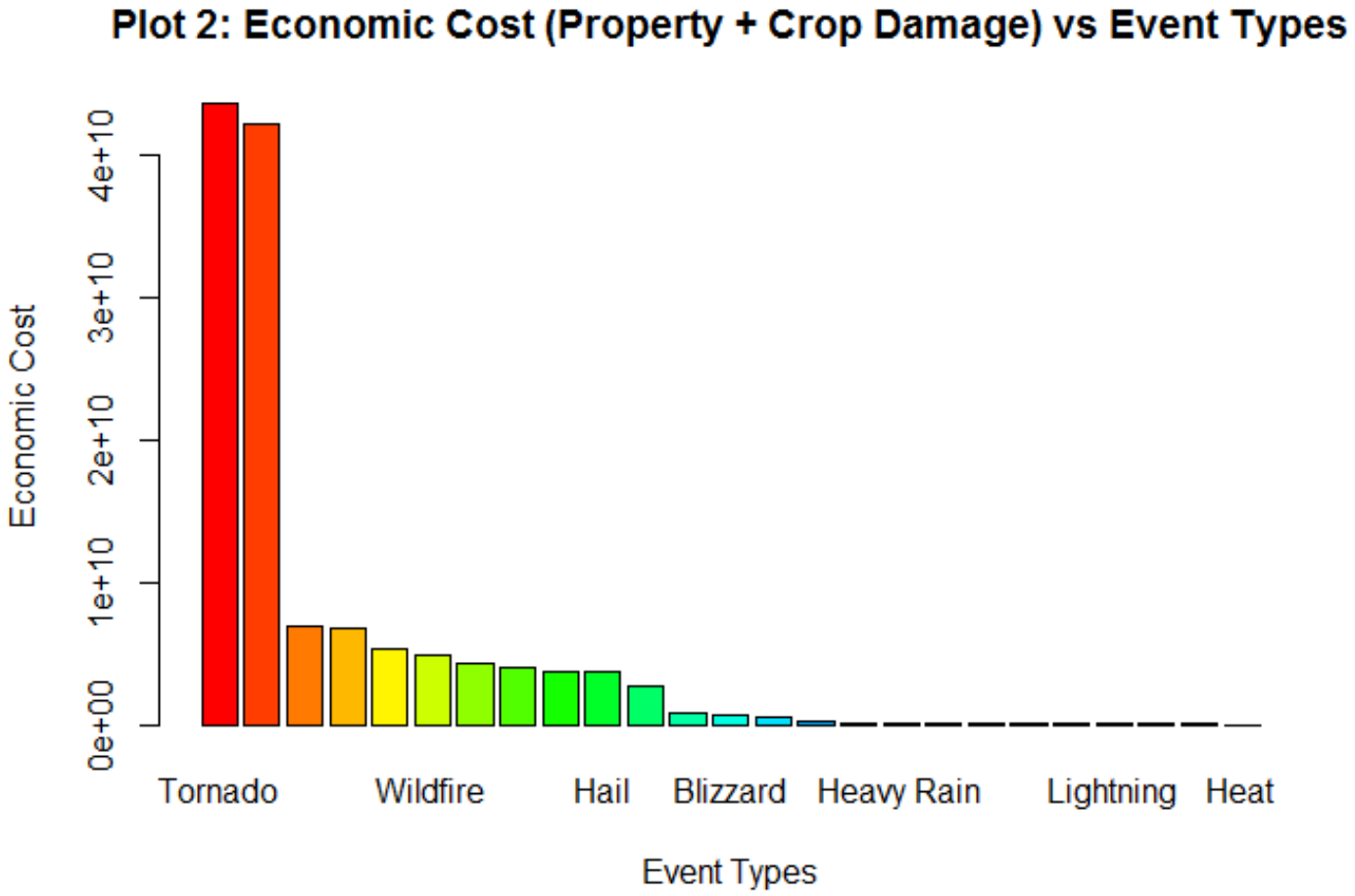
Table 1: Health Impact for top 25 event types

```
##---End of Code to Answer Q1 ---
```

- The maximum health impact is due to event **Tornado** in terms of total human fatalities and injuries of **97,022**

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

```
        ##---Start of Code to Answer Q2 ---
        ##Using storm data frame from loaddata r code chunk above. Remove the NAs. Retain
values that are non-zero.  This is being done given the question to be answered w.r.t pro
perty damage and crop damage (economic cost)
        stormdata<-subset(stormdata, !is.na(PROPDMG) & !is.na(CROPDMG) & (PROPDMG>0 | CRO
PDMG>0))



        ##Convert the Property Damage and Crop Damage Cost using the repaired exponents
        stormdata <- mutate(stormdata, PROPDMG=PROPDMG*(10^(as.numeric(PROPDMGEXP))), CRO
PDMG=CROPDMG*(10^(as.numeric(CROPDMGEXP))))

        ##One can use group by approach below but I like melt and cast
        ##Now melt the data set with event type as ID.
        meltedstormdatabyevent <- melt(stormdata, id.vars = c("EVTYPE"), measure.var
s=c("PROPDMG", "CROPDMG"))

        ##Then cast the data set with event type as ID and calculate the average by event
type
        castedstormdatabyevent <- dcast(meltedstormdatabyevent, EVTYPE~variable,sum)

        ##Introduce a column ECONOMICCOST that is a sum of PROPDMG and CROPDMG
        castedstormdatabyevent <- mutate(castedstormdatabyevent, ECONOMICCOST=round(PROPD
MG+CROPDMG))

        ##Order by ECONOMICCOST
        castedstormdatabyevent<-arrange(castedstormdatabyevent,desc(ECONOMICCOST))

        ##Get the max economic cost and the corresponding event
        maxeconomiccost<-max(castedstormdatabyevent$ECONOMICCOST)
        eventtypewithmaxeconomiccost<-castedstormdatabyevent[which(castedstormdatabyeven
t$ECONOMICCOST==maxeconomiccost), ][1,1]

        ##Plot and print the table for Top 25 events only
        top25<-head(castedstormdatabyevent, 25)
        par(mfcol=c(1,1))
        barplot(top25$ECONOMICCOST, names.arg=top25$EVTYPE, main="Plot 2: Economic Cost
(Property + Crop Damage) vs Event Types",
          xlab="Event Types", ylab="Economic Cost", col=rainbow(25))
```




```
        economiccost<-xtable(top25, caption="Table 2: Economic Cost for top 25 event type
s")
        print(economiccost, type="html")
```

	EVTYPE	PROPDMG	CROPDMG	ECONOMICCOST
1	Tornado	43487203105.00	143977000.00	43631180105.00
2	Hurricane/Typhoon	38314160000.00	3825350800.00	42139510800.00
3	Flood	6758761640.00	169075500.00	6927837140.00
4	Tropical Storm	6560656000.00	157765000.00	6718421000.00
5	Winter Storm	5281662500.00	15783000.00	5297445500.00
6	Wildfire	4667112200.00	188627000.00	4855739200.00
7	Flash Flood	4229270800.00	48871750.00	4278142550.00
8	Storm Surge/Tide	4066020000.00	0.00	4066020000.00
9	High Wind	3329602140.00	354630150.00	3684232290.00
10	Hail	3505785701.00	166300000.00	3672085701.00
11	Thunderstorm Wind	2491423793.00	184893550.00	2676317343.00
12	Ice Storm	772861000.00	20260000.00	793121000.00
13	Blizzard	526756000.00	112050000.00	638806000.00
14	Excessive Heat	3607500.00	492400000.00	496007500.00
15	Heavy Snow	225383250.00	170000.00	225553250.00

16	Strong Wind	82985800.00	63450000.00	146435800.00
17	Heavy Rain	109939000.00	25000000.00	134939000.00
18	Tsunami	84000000.00	20000.00	84020000.00
19	Waterspout	50110500.00	0.00	50110500.00
20	Frost/Freeze	0.00	50000000.00	50000000.00
21	Extreme Cold/Wind Chill	26037000.00	1750000.00	27787000.00
22	Lightning	22335783.00	3510050.00	25845833.00
23	Extreme Cold	0.00	20000000.00	20000000.00
24	Dense Fog	16480000.00	0.00	16480000.00
25	Heat	8415050.00	1535000.00	9950050.00

Table 2: Economic Cost for top 25 event types

```
##---End of Code to Answer Q2 ---
```

- The maximum economic cost is due to event **Tornado** with a total cost of **43,631,180,105**

————————————————End of the report————————————————