Population Health and Economic Impacts of Storm Events

Simon Cox

30 Nov 2019

Synopsis

The following document provides an analysis of storm data from the US National Oceanic and Atmospheric Administration's storm database, and its impact on human population health and the economy.

In summary the most significant types of storm events are:

- tornado (for human fatalities / injuries)
- flood (for property damage)
- drought (for crop damage).

The results show quite strongly that tornadoes are the most significant type of storm event when it comes to population health, causing the most human injuries and fatalities. However it is interesting to note that the most damaging types of storm events cross a range of conditions including wind-related (tornadoes, TSTM Wind), heat-related (excessive heat and heat), water-related (flood and flash flood) and other types of events (e.g. lightning and ice storm).

With regards to the economic impacts of storm events, 5 of the top 6 types of storm events causing property damage are all water-related (flood, hurricane/typhoon, storm surge, flash flood, hail), with tornadoes another major contributor.

However when it comes to the types of events that cause the most signficant crop damage we see that drought is by far the most significant, along with flood, river flood, ice storm, hail and hurricane / typhoon. Interestingly, tornadoes are not a big culprit for causing crop damage, sitting at number 17 on the list. This is possibly due to the fact that farmers choose not to farm in areas that are susceptible to tornado activity.

Data Processing

The raw data is taken from the U.S. National Oceanic and Atmospheric Administration's (NOAA) storm database, available here. This data is compressed using the bzip2 algorithm, and so can be read directly into R using the standard read.csv function. The code used is shown below:

```
# download file to ./data directory if it doesn't already exist
if (!file.exists("data")) {
    dir.create("data")
}

if (!file.exists("./data/storm_data.csv.bz2")) {
    fileURL <- "https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2"
    download.file(fileURL, destfile = "./data/storm_data.csv.bz2")
}

# load data into dataframe
storm_raw <- read.csv("./data/storm_data.csv.bz2",stringsAsFactors = FALSE)
```

Data Transformations

An initial analysis of the data showed that there are a small number of columns of interest in order to understand the impact of storm events on population health:

- EVTYPE - which defines the storm event type
- FATALITIES - the number of human fatalities attributable to the storm event
- INJURIES - the number of human injuries attributable to the storm event

Additionally there are a small number of columns of interest in order to understand the economic impact of storm events:

- EVTYPE - which defines the storm event type
- PROPDMG - a number providing a representation of the cost of the property damage
- PROPEXP - a value that defines the multiplication factor against PROPDMG to get the value of the damage
- CROPDMG - a number providing a representation of the cost of the crop damage
- CROPEXP - a value that defines the multiplication factor against CROPDMG to get the value of the damage

The actual value of property / crop damage is not provided in the data. Therefore some transformation is required. A very helpful individual (David Song) left a comment on the project discussion board which provided a decoder for the PROPEXP and CROPEXP fields at this link.
Using this decoder the true value of the damage to property and crops can be calculated in R:

```r
# reduce data to information of interest
storm <- storm_raw %>%
    select(EVTYPE,
           FATALITIES,
           INJURIES,
           PROPDMG,
           PROPDMGEXP,
           CROPDMG,
           CROPDMGEXP)

# create multiplier for exponent data
mult <- data.frame(exp = c("","-","?","+",0:8,"B","h","H","k","K","m","M")
                   , MULT = c(0,0,0,rep(1,10),1000000000,100,100,1000,1000,1000000,1000000))

# merge multiplier values into storm dataframe for crop damage and property damage
storm <- merge(storm,mult,by.x = "PROPDMGEXP", by.y = "exp")
storm <- storm %>% rename(PROPDMGMULT = MULT)
storm <- merge(storm,mult,by.x = "CROPDMGEXP", by.y = "exp")
storm <- storm %>% rename(CROPDMGMULT = MULT)

# create new columns for value of property damage and crop damage
storm <- storm %>%
```

```
    mutate(PROPDMGVAL = PROPDMG * PROPDMGMULT) %>%
    mutate(CROPDMGVAL = CROPDMG * CROPDMGMULT) %>%
    select(EVTYPE,
           FATALITIES,
           INJURIES,
           PROPDMGVAL,
           CROPDMGVAL)

# count the number of unique Event Types
unique_evtype <- length(unique(storm$EVTYPE))
```

Data Analysis - Event Types

It is important to understand the data contained in the Event Type field of the dataset, in order to work out if any transformation is required. The documentation for the dataset suggests that there are around 40 unique event types that should have been logged in the database, so we should expect to see this in the actual data.
A quick calculation shows that there are **985** unique values in the event type field, which is significantly more than expected most likely due to human input of character strings into the database rather than provision of a categorical dropdown input.
Therefore some further analysis of this field is required - specifically are there a number of event types that have been commonly used and a small number that have only been used relatively rarely?
The top 100 event types were determined and are output below:

```
top_100 <- as.data.frame(storm %>%
                            group_by(EVTYPE) %>%
                            summarise(nbr = n()) %>%
                            top_n(100,nbr) %>%
                            arrange(desc(nbr)))
names(top_100) <- c("Event Type","Number of Events")
kable(top_100)
```

Event Type	Number of Events
HAIL	288661
TSTM WIND	219940
THUNDERSTORM WIND	82563
TORNADO	60652
FLASH FLOOD	54277
FLOOD	25326
THUNDERSTORM WINDS	20843
HIGH WIND	20212
LIGHTNING	15754
HEAVY SNOW	15708
HEAVY RAIN	11723
WINTER STORM	11433
WINTER WEATHER	7026
FUNNEL CLOUD	6839
MARINE TSTM WIND	6175
MARINE THUNDERSTORM WIND	5812
WATERSPOUT	3796
STRONG WIND	3566

Event Type	Number of Events
URBAN/SML STREAM FLD	3392
WILDFIRE	2761
BLIZZARD	2719
DROUGHT	2488
ICE STORM	2006
EXCESSIVE HEAT	1678
HIGH WINDS	1533
WILD/FOREST FIRE	1457
FROST/FREEZE	1342
DENSE FOG	1293
WINTER WEATHER/MIX	1104
TSTM WIND/HAIL	1028
EXTREME COLD/WIND CHILL	1002
HEAT	767
HIGH SURF	725
TROPICAL STORM	690
FLASH FLOODING	682
EXTREME COLD	655
COASTAL FLOOD	650
LAKE-EFFECT SNOW	636
FLOOD/FLASH FLOOD	624
LANDSLIDE	600
SNOW	587
COLD/WIND CHILL	539
FOG	538
RIP CURRENT	470
MARINE HAIL	442
DUST STORM	427
AVALANCHE	386
WIND	340
RIP CURRENTS	304
STORM SURGE	261
FREEZING RAIN	250
URBAN FLOOD	249
HEAVY SURF/HIGH SURF	228
EXTREME WINDCHILL	204
STRONG WINDS	196
DRY MICROBURST	186
ASTRONOMICAL LOW TIDE	174
HURRICANE	174
RIVER FLOOD	173
LIGHT SNOW	154
STORM SURGE/TIDE	148
RECORD WARMTH	146
COASTAL FLOODING	143
DUST DEVIL	141
MARINE HIGH WIND	135
UNSEASONABLY WARM	126
FLOODING	120
ASTRONOMICAL HIGH TIDE	103
MODERATE SNOWFALL	101
URBAN FLOODING	98

Event Type	Number of Events
WINTRY MIX	90
HURRICANE/TYPHOON	88
FUNNEL CLOUDS	87
HEAVY SURF	84
RECORD HEAT	81
FREEZE	74
HEAT WAVE	74
COLD	72
RECORD COLD	64
ICE	61
THUNDERSTORM WINDS HAIL	61
TROPICAL DEPRESSION	60
SLEET	59
UNSEASONABLY DRY	56
FROST	53
GUSTY WINDS	53
THUNDERSTORM WINDSS	51
MARINE STRONG WIND	48
OTHER	48
SMALL HAIL	47
FUNNEL	46
FREEZING FOG	45
THUNDERSTORM	45
Temperature record	43
TSTM WIND (G45)	39
Coastal Flooding	38
WATERSPOUTS	37
MONTHLY PRECIPITATION	36
WINDS	36
MIXED PRECIPITATION	34

This shows us that out of the total number of unique event types, only 69 have more than 100 entries in the dataset.
Ideally we'd run a detailed data cleanse activity to try and fit all of the unique event types into the standard 40 event types suggested by the dataset documentation but time constraints prevent us from doing so. Therefore our analysis will make use of the event types as entered directly into the database, recognising that this is not ideal.

Results

Population Health Impacts

The health impacts are described by 2 fields - the number of injuries and the number of fatalities, and we can look at the storm events that impact on these in aggregate.

```
by_event <- storm %>%
    mutate(INCIDENTS = INJURIES + FATALITIES) %>%
    group_by(EVTYPE)
incidents <- by_event %>%
    summarise(Total_Incidents = sum(INCIDENTS)) %>%
```

```r
    arrange(desc(Total_Incidents)) %>%
    top_n(20,Total_Incidents)

g <- ggplot(data = incidents, aes(x = EVTYPE
                                  , y = Total_Incidents
                                  , fill=EVTYPE
                                  , label = Total_Incidents)) +
    geom_bar(stat="identity"
             , aes(reorder(EVTYPE,Total_Incidents),Total_Incidents)) +
    geom_text(size = 2.5
              , aes(label = format(Total_Incidents, big.mark = ","))
              , position = position_stack(vjust = 0.5, reverse = FALSE)) +
    scale_y_continuous(labels = comma) +
    coord_flip() +
    labs(title = "Fatalities / Injuries by Event Type"
         , caption = "Figure 1 - Impact on population health from storm events"
         , y = "Number of Incidents"
         , x = "Event Type") +
    theme(legend.position = ""
          , axis.text = element_text(size = 8)
          , plot.caption = element_text(hjust = 0,face = "italic")
          , plot.title = element_text(hjust = 0.5, size = 15))

print(g)
```

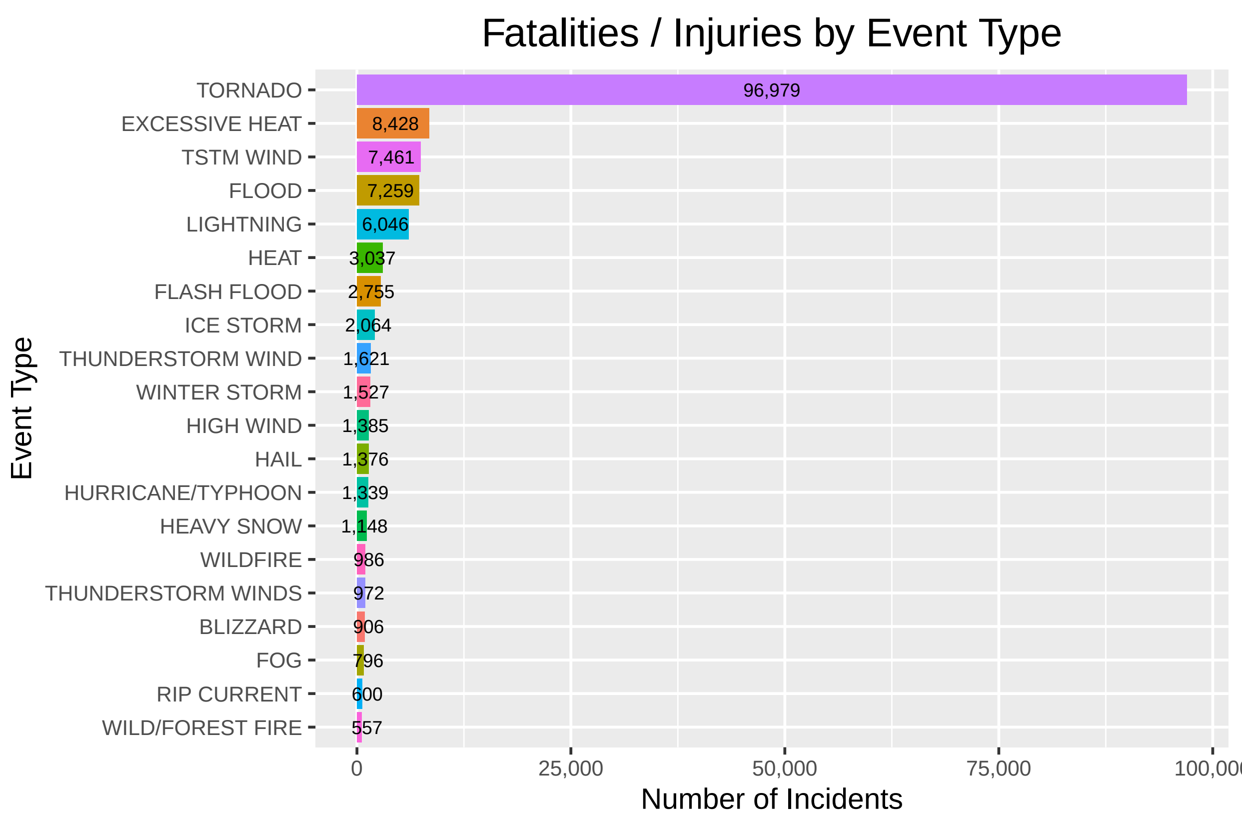

Figure 1 − Impact on population health from storm events

The chart shows quite conclusively that by far the most detrimental storm event to the health of the population is the tornado. A list of the most notable types of storm event that cause significant injuries and fatalities is below:

1. Tornado

2. Excessive Heat

3. TSTM Wind
4. Flood
5. Lightning

Economic Impacts

The economic impacts are described by 2 fields - the value of property damage and the value of crop damage, and we can look at the storm events that impact on each of these in turn.

```
usd <- dollar_format(largest_with_cents = 0, suffix = "m")
propdmg <- by_event %>%
    mutate(Property_Damage_m = PROPDMGVAL / 1000000) %>%
    summarise(Property_Damage = sum(Property_Damage_m)) %>%
    arrange(desc(Property_Damage)) %>%
    top_n(20,Property_Damage)

g <- ggplot(data = propdmg, aes(x = EVTYPE
                                , y = Property_Damage
                                , fill=EVTYPE
                                , label = Property_Damage)) +
    geom_bar(stat="identity"
            , position = "dodge"
            , aes(reorder(EVTYPE,Property_Damage),Property_Damage)) +
    geom_text(size = 2.5
             , position = position_stack(vjust = 0.5)
             , aes(label = usd(Property_Damage))) +
    scale_y_continuous(breaks = c(seq(0,140000,20000)), labels = comma) +
    coord_flip() +
    labs(title = "Value of Property Damage by Event Type"
         , caption = "Figure 2 - Economic Impact on Property from storm events"
         , y = "Property Damage ($million)"
         , x = "Event Type") +
    theme(legend.position = ""
          , axis.text = element_text(size = 8)
          , plot.caption = element_text(hjust = 0, face = "italic")
          , plot.title = element_text(hjust = 0.5, size = 15))

print(g)
```

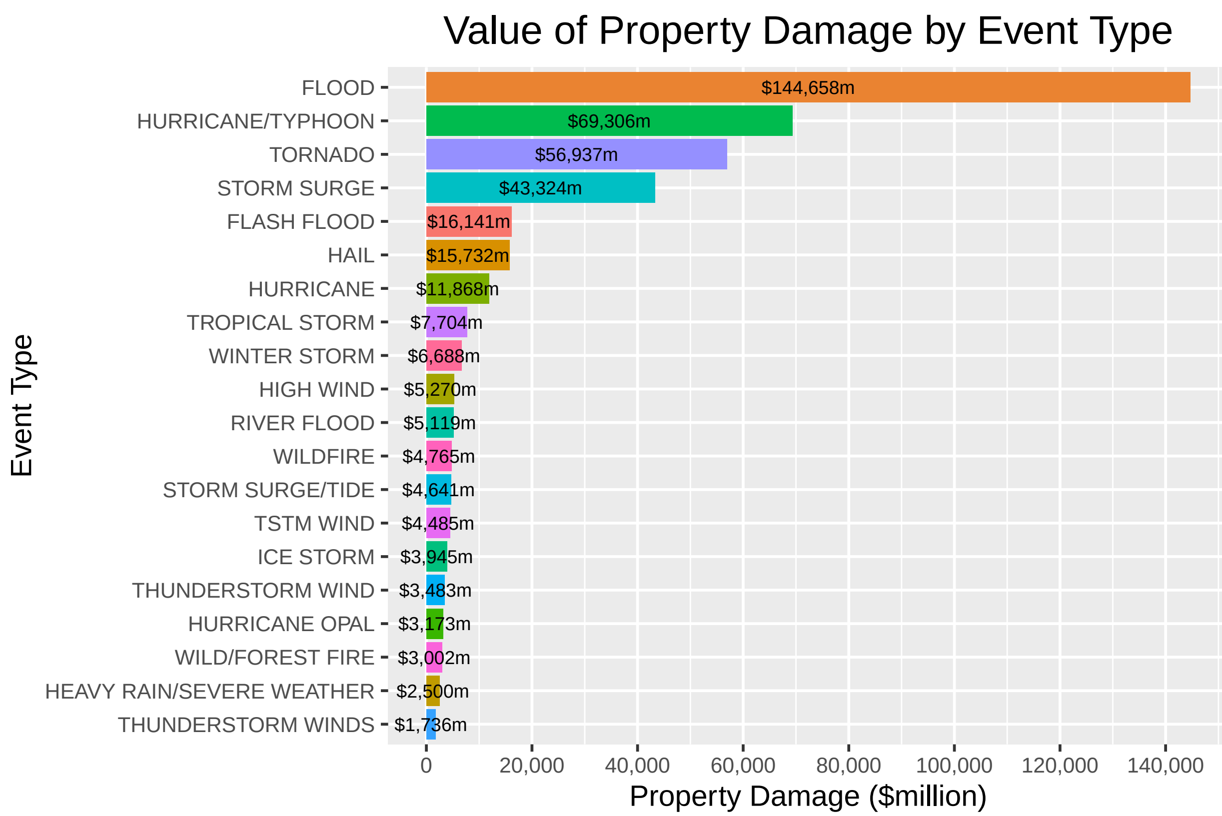

Figure 2 − Economic Impact on Property from storm events

Figure 2 shows that with the exception of tornadoes, water-related events cause the most damage to property, with the most significant types of storm events being:

1. Flood
2. Hurricane / Typhoon
3. Tornado
4. Storm Surge
5. Flash Flood

```
usd <- dollar_format(largest_with_cents = 0, suffix = "m")
cropdmg <- by_event %>%
    mutate(Crop_Damage_m = CROPDMGVAL / 1000000) %>%
    summarise(Crop_Damage = sum(Crop_Damage_m)) %>%
    arrange(desc(Crop_Damage)) %>%
    top_n(20,Crop_Damage)

g <- ggplot(data = cropdmg, aes(x = EVTYPE
                                , y = Crop_Damage
                                , fill=EVTYPE
                                , label = Crop_Damage)) +
    geom_bar(stat="identity"
             , position = "dodge"
```

```r
          , aes(reorder(EVTYPE,Crop_Damage),Crop_Damage)) +
  geom_text(size = 2.5
            , position = position_stack(vjust = 0.5)
            , aes(label = usd(Crop_Damage))) +
  scale_y_continuous(breaks = c(seq(0,14000,2000)), labels = comma) +
  coord_flip() +
  labs(title = "Value of Crop Damage by Event Type"
       , caption = "Figure 3 - Economic Impact on Crops fom storm events"
       , y = "Crop Damage ($million)"
       , x = "Event Type") +
  theme(legend.position = ""
        , axis.text = element_text(size = 8)
        , plot.caption = element_text(hjust = 0, face = "italic")
        , plot.title = element_text(hjust = 0.5, size = 15))

print(g)
```

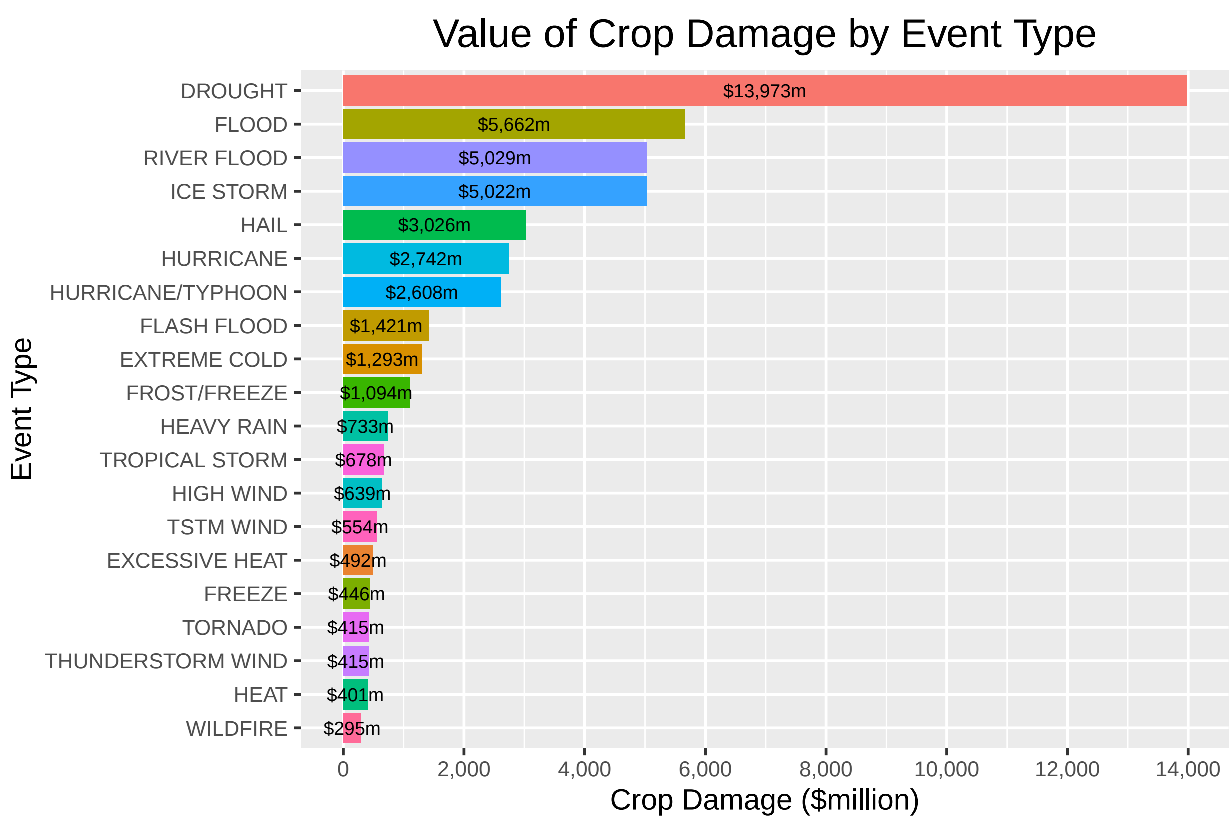

Figure 3 − Economic Impact on Crops fom storm events

Figure 3 shows more of a mix of events that cause damage to crops. However it is very clear that the most damaging event is drought which causes more than 2.5 times the damage than the next item which is flood. The most significant types of storm events that cause crop damage are:

1. Drought
2. Flood

3. River Flood

4. Ice Storm

5. Hail

Figures 2 and 3 show that the total value of property damage is much higher than the total value of crop damage, by about a factor of 10.