Analysis of Storm Events and their effect on Economy and Health

Sahil

2/1/2021

Synopsis

U.S. National Oceanic and Atmospheric Administration's (NOAA) storm database has been tracking a standard set of 48 storm data events since 1996. The purpose of this analysis was to determine which storm events have the most significant effectw on economy and health of populations. Upon analysis of the provided storm data event it was found that Hurricanes & Typhoons cause the most economic damage via crop and property damage, while Tornadoes are responsible for the highest number of injuries and fatalities attributable to a single type of storm event.

Data Processing

Retrieval and Loading

The compressed data is conditionally downloaded from the source URL, if not found locally, and then loaded via the function `read.csv`.

```
if (!file.exists('StormData.csv.bz2')) {
  download.file('https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2', 'StormData.
}
storm_data <- read.csv('StormData.csv.bz2')

# Ensure data is downloaded, decompressed and loaded correctly by checking filesize and dimemsions
stopifnot(dim(storm_data) == c(902297,37))
```

Pre-Processing

Magnitude (Dollar value) of Property Damage and Crop Damage variables was noted by an "EXP" value to indicate how much to multiply the DMG vairable by to arrive at the total value. Ex. If PROPDMG = 25 and PROPEXP is "K", it is supposed to indicate a total of $ 25,000 in Damage. TO run calculations we will need to convert the EXP values to their appropriate numerica value, and multiply by the corresponding DMG value. The following code chunk are used for this purpose

```
convEXPtoMult <- function(exp) {
  ifelse(
    exp == '+', 1,                     # '+' -> 1
    ifelse(
      exp %in% paste(seq(0,8)), 10^1,  # 0-8 -> 10
      ifelse(
        exp %in% c('H', 'h'), 10^2,    # H,h -> 100
```

```
      ifelse(
        exp %in% c('K', 'k'), 10^3,     # K,k -> 1,000
        ifelse(
          exp %in% c('M', 'm'), 10^6,    # M,m -> 1,000,000
          ifelse(
            exp %in% c('B', 'b'), 10^9,  # B,b -> 1,000,000,000
            0                             # everything else -> 0
          )
        )
      )
    )
  )
}
storm_data$PropDamageMult <- convEXPtoMult(storm_data$PROPDMGEXP)
storm_data$CropDamageMult <- convEXPtoMult(storm_data$CROPDMGEXP)
```

now we multiply the columns to get the dollar value and store the result in new columns - PropDamageValue & CropDamageValue. We Add third column called TotalDamage which is the sum of the former two columns mentioned.

```
storm_data$PropDamageValue  <- storm_data$PROPDMG * storm_data$PropDamageMult
storm_data$CropDamageValue  <- storm_data$CROPDMG * storm_data$CropDamageMult
storm_data$TotalDamage <- storm_data$PropDamageValue + storm_data$CropDamageValue
```

To determine the overall health effect of storm events, a `TotalHealthImpact` variable is added using the sum of `FATALITIES` and `INJURIES` variables.

```
storm_data$TotalHealthImpact <- storm_data$FATALITIES + storm_data$INJURIES
```

The years from the dataset earlier than 1996 that are trimmed in order to allow for a fair assessment of variables in question. Also, since we are answering questions regarding economic and health impact, all rows where value is 0 for both damage and health impact should be removed

```
storm_data$BeginDate    <- as.Date(storm_data$BGN_DATE, '%m/%d/%Y')
sd <- storm_data[storm_data$BeginDate >= '1996-01-01',]
sd <- sd[sd$TotalDamage > 0 | sd$TotalHealthImpact  > 0,]
```

5. Looking at the top events with the most `TotalDamage` and `TotalHealthImpact` showed that there was a least one event that had far more economic impact than any other. Using the NOAA Storm Events Database, it was found that a 2006 flood in Napa County, Califorina was mis-entered with a `PROPDMGEXP` of **B**illion instead of **M**illion.

The erroneous `PROPDMGEXP` value was then corrected and the `PropDamageMult`, `PropDamageValue` and `TotalDamage` variables were recalculated.

```
sd$PROPDMGEXP[sd$REFNUM=='605943'] <- 'M'
sd$PropDamageMult <- convEXPtoMult(sd$PROPDMGEXP)
sd$PropDamageValue  <- sd$PROPDMG * sd$PropDamageMult
sd$TotalDamage <- sd$PropDamageValue + sd$CropDamageValue
```

After checking the remaining top 5 by damage and health impact, it was found those are consistent with data available in the NOAA database. See appendix B for more information on the checks of the top individual events.

6. EVTYPE column does not use consistent naming of events

The `EVTYPE` variable was first updated for consistency by removing all whitespace and making all upper case.

```
sd$EVTYPE <- toupper(trimws(sd$EVTYPE))
```

EVTYPE values were made more consistent with the following code

```
coastalFloodAliases <- c('ASTRONOMICAL HIGH TIDE', 'TIDAL FLOODING', 'COASTAL FLOODING/EROSION', 'COAST
                         'COASTAL  FLOODING/EROSION', 'EROSION/CSTL FLOOD')
sd$EventType[sd$EVTYPE %in% coastalFloodAliases] <- 'Coastal Flood'

winterWeatherAliases <- c('LIGHT FREEZING RAIN', 'ICY ROADS', 'GLAZE', 'FREEZING RAIN',
                          'FREEZING DRIZZLE', 'LIGHT SNOW', 'LIGHT SNOWFALL', 'WINTER WEATHER/MIX',
                          'MIXED PRECIPITATION', 'MIXED PRECIP', 'WINTRY MIX', 'RAIN/SNOW',
                          'WINTER WEATHER MIX')
sd$EventType[sd$EVTYPE %in% winterWeatherAliases] <- 'Winter Weather'
heavySnowAliases <- c('EXCESSIVE SNOW', 'AVALANCHE', 'SNOW', 'HEAVY SNOW SHOWER', 'SNOW SQUALL', 'SNOW S
sd$EventType[sd$EVTYPE %in% heavySnowAliases] <- 'Heavy Snow'
highWindAliases <- c('WIND', 'WINDS', 'GUSTY WINDS', 'GUSTY WIND', 'HIGH WIND (G40)', 'HIGH WIND',
                     'NON TSTM WIND',  'NON-TSTM WIND', 'WIND DAMAGE', 'NON TSTM WIND',
                     'NON-SEVERE WIND DAMAGE', 'GRADIENT WIND', 'STRONG WIND', 'STRONG WINDS' )
sd$EventType[sd$EVTYPE %in% highWindAliases] <- 'High Wind'
freezeAliases <- c('FREEZE', 'DAMAGING FREEZE', 'EARLY FROST', 'FROST', 'AGRICULTURAL FREEZE',
                   'HARD FREEZE', 'UNSEASONABLY COLD', 'UNSEASONABLE COLD')
sd$EventType[sd$EVTYPE %in% freezeAliases] <- 'Frost/Freeze'
extremeColdAliases <- c('EXTREME WINDCHILL', 'EXTREME COLD')
sd$EventType[sd$EVTYPE %in% extremeColdAliases] <- 'Extreme Cold/Wind Chill'
floodAliases <- c('RIVER FLOODING', 'RIVER FLOOD', 'URBAN/SML STREAM FLD', 'URBAN FLOOD')
sd$EventType[sd$EVTYPE %in% floodAliases] <- 'Flood'
flashFloodAliases <- c('FLASH FLOOD/FLOOD', 'FLOOD/FLASH/FLOOD')
sd$EventType[sd$EVTYPE %in% flashFloodAliases] <- 'Flood'
thunderstormAliases <- c('TSTM WIND', 'TSTM WINDS', 'THUNDERSTORM', 'THUNDERSTORMS', 'THUNDERSTORM WIND
                         'THUNDERSTORM WINDSS', 'THUNDERSTORMS WINDS', 'DRY MICROBURST',
                         'TSTM WIND (G40)', 'THUNDERSTORM WIND/ TREES', 'MICROBURST',
                         'WET MICROBURST', 'THUNDERTORM WINDS', 'THUNDERSTORMS WIND',
                         'SEVERE THUNDERSTORM WINDS', 'TSTM WIND 55', 'THUNDERSTORM WIND 60 MPH',
                         'TSTM WIND (G45)', 'SEVERE THUNDERSTORM', 'THUDERSTORM WINDS',
                         'THUNDEERSTORM WINDS', 'THUNDERESTORM WIND', 'TSTM WIND 40',
                         'TSTM WIND G45', 'TSTM WIND  (G45)', 'TSTM WIND (41)', 'TSTM WIND 45',
                         'TSTM WIND (G35)', 'TSTM WIND AND LIGHTNING', 'TSTM WIND/HAIL',
                         'THUNDERSTORM WIND (G40)', 'LIGHTNING')
sd$EventType[sd$EVTYPE %in% thunderstormAliases] <- 'Thunderstorm Wind'
hailAliases <- c('HAIL DAMAGE', 'SMALL HAIL', 'HAILSTORM')
sd$EventType[sd$EVTYPE %in% hailAliases] <- 'Hail'
hurricaneAliases <- c('HURRICANE', 'TYPHOON', 'HURRICANE OPAL', 'HURRICANE ERIN',
                      'HURRICANE EDOUARD', 'HURRICANE EMILY', 'HURRICANE FELIX',
                      'HURRICANE GORDON', 'HURRICANE OPAL/HIGH WINDS')
sd$EventType[sd$EVTYPE %in% hurricaneAliases] <- 'Hurricane (Typhoon)'
```

```
highSurfAliases <- c('HEAVY SURF/HIGH SURF', 'HEAVY SURF', 'HIGH SURF ADVISORY')
sd$EventType[sd$EVTYPE %in% highSurfAliases] <- 'High Surf'
wildfireAliases = c('WILD/FOREST FIRE', 'BRUSH FIRE')
sd$EventType[sd$EVTYPE %in% wildfireAliases] <- 'Wildfire'
heatAliases = c('UNSEASONABLY WARM', 'WARM WEATHER')
sd$EventType[sd$EVTYPE %in% heatAliases] <- 'Heat'
excessiveHeatAliases = c('HEAT WAVE', 'RECORD HEAT')
sd$EventType[sd$EVTYPE %in% excessiveHeatAliases] <- 'Excessive Heat'
heavyRainAliases = c('TORRENTIAL RAINFALL', 'RAIN', 'UNSEASONAL RAIN')
sd$EventType[sd$EVTYPE %in% heavyRainAliases]  <- 'Heavy Rain'
# one-offs
sd$EventType[sd$EVTYPE == 'TORNADO']          <- 'Tornado'
sd$EventType[sd$EVTYPE == 'LANDSPOUT']        <- 'Tornado'
sd$EventType[sd$EVTYPE == 'FOG']              <- 'Dense Fog'
sd$EventType[sd$EVTYPE == 'MARINE TSTM WIND'] <- 'Marine Thunderstorm Wind'
sd$EventType[sd$EVTYPE == 'LANDSLIDE']        <- 'Debris Flow'
sd$EventType[sd$EVTYPE == 'STORM SURGE']      <- 'Storm Surge/Tide'
sd$EventType[sd$EVTYPE == 'COLD']             <- 'Cold/Wind Chill'
```

Results

Event Types Most Harmful to Population Health

```
top_health <- head(
  arrange(
    aggregate(
      cbind(FATALITIES, INJURIES, TotalHealthImpact) ~ EVTYPE, sd, FUN = sum),
    desc(TotalHealthImpact)
  ),
  n=5
)
kable(
  top_health,
  caption = 'Fig 1. Top 5 Event Types Most Harmful to Population Health'
)
```

Table 1: Fig 1. Top 5 Event Types Most Harmful to Population Health

EVTYPE	FATALITIES	INJURIES	TotalHealthImpact
TORNADO	1511	20667	22178
EXCESSIVE HEAT	1797	6391	8188
FLOOD	414	6758	7172
LIGHTNING	651	4141	4792
TSTM WIND	241	3629	3870

Fig 1 Explanation:

Tornado events top the list with more than double the health impact of second place, which is Excessive Heat.

Although Excessive Heat's total health impact score is far below tornadoes it has the most fatalities overall, making it the most lethal of all events since 1996.

```
tophealthplot <- ggplot(top_health, aes(x= EVTYPE, y= TotalHealthImpact, fill = EVTYPE)) + geom_bar(sta
tophealthplot
```
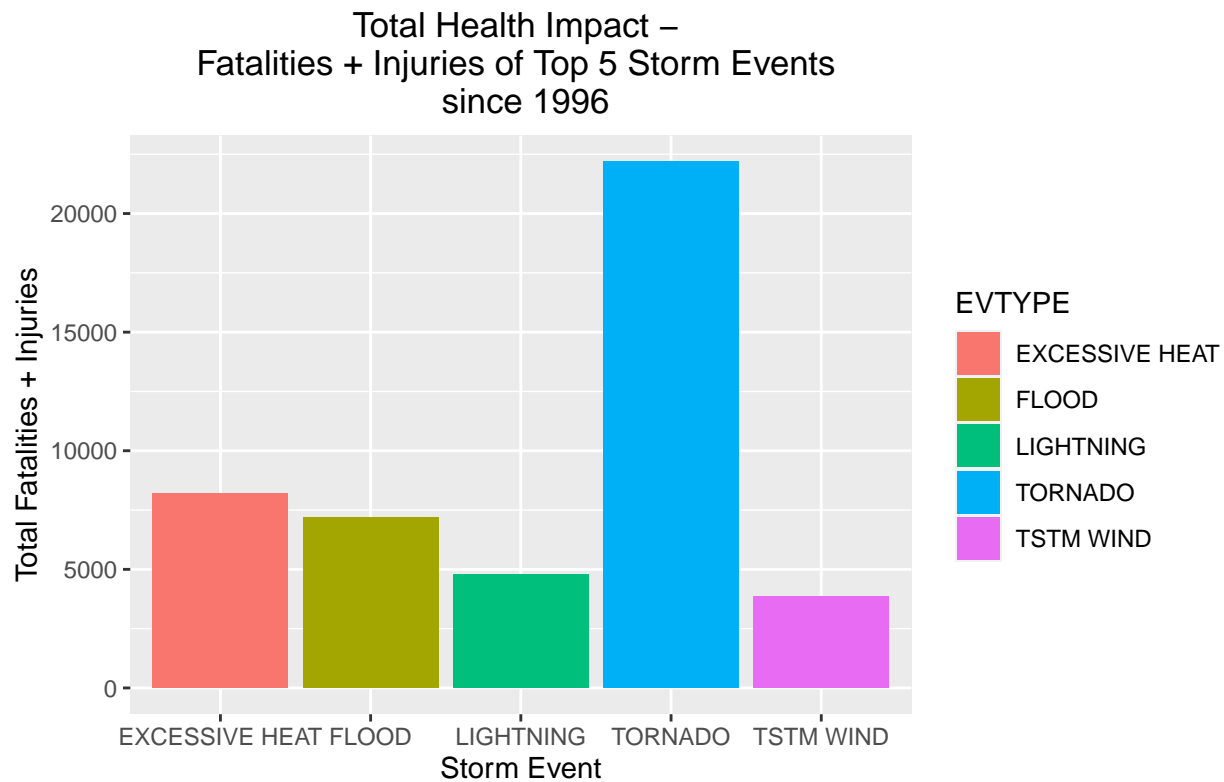

Fig. 2, Total Health Impact
for each of the top 5 most damaging events

Fig 2 Explanation:

Here we see a graphical representation of the events with the greatest impact on public health. Tornadoes total health impact is upwards of 20,000 since 1996.

Event Types with Greatest Economic Consequences

```
top_damage <- head(
  arrange(
    aggregate(
      cbind(CropDamageValue, PropDamageValue, TotalDamage) ~ EVTYPE, sd, FUN=sum),
    desc(TotalDamage)
  ),
  n=5
)
#kable(
  #top_damage,
  #format.args = list(big.mark = ","),
  #caption = 'Fig. 3 Top 5 Event Types with Greatest Economic Consequences')
```

```
topdmgplot <- ggplot(top_damage, aes(x= EVTYPE, y= TotalDamage, fill = EVTYPE)) + geom_bar(stat='identi
```
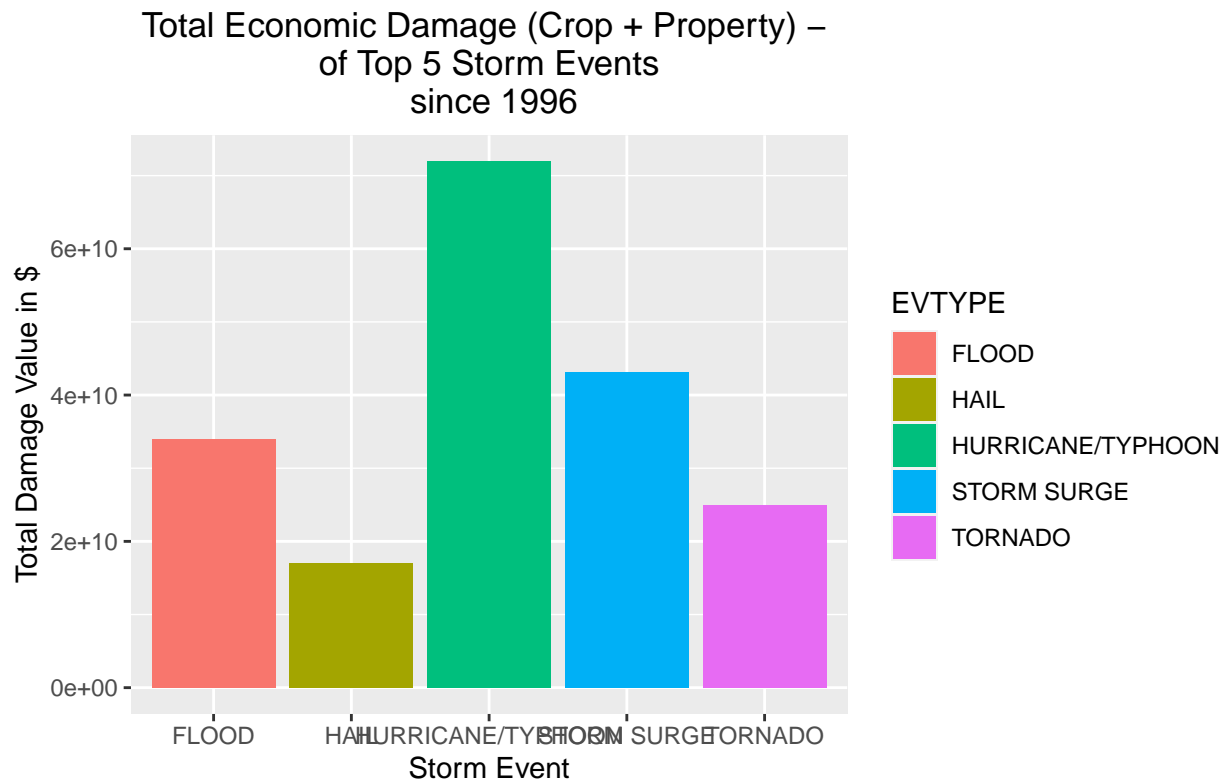
```
topdmgplot
```


Fig. 3, Total value of economic damage
for each of the top 5 most damaging events

Fig 3. Explanation:

Here we see that Hurricane/Typhoon events have caused the most damage at ~$70+ billion

Storm Surge/Tide events come in a close second at ~$40+ billion.

Conclusion

Thus we see that storm events have a major impact on population health and local economies. Tornadoes cause the most injuries and fatalities. Hurricane type events cause the most economic damage.