US Storm Analysis

Dinh Tuan Phan

4/29/2021

Synopsis

Storms and other severe weather events can cause both public health and economic problems for communities and municipalities. Many severe events can result in fatalities, injuries, and property damage, and preventing such outcomes to the extent possible is a key concern.

This project involves exploring the U.S. National Oceanic and Atmospheric Administration's (NOAA) storm database. This database tracks characteristics of major storms and weather events in the United States, including when and where they occur, as well as estimates of any fatalities, injuries, and property damage.

Data Processing

Load packages and data

```
library("data.table")
```

```
## Warning: package 'data.table' was built under R version 3.6.3
```

```
library("ggplot2")

stormDT <- read.csv('repdata_data_StormData.csv')
stormDT <- as.data.table(stormDT)
```

First look at data

```
summary(stormDT)
```

```
##     STATE__                  BGN_DATE             BGN_TIME
##  Min.   : 1.0   5/25/2011 0:00:00:  1202   12:00:00 AM: 10163
##  1st Qu.:19.0   4/27/2011 0:00:00:  1193   06:00:00 PM:  7350
##  Median :30.0   6/9/2011 0:00:00 :  1030   04:00:00 PM:  7261
##  Mean   :31.2   5/30/2004 0:00:00:  1016   05:00:00 PM:  6891
##  3rd Qu.:45.0   4/4/2011 0:00:00 :  1009   12:00:00 PM:  6703
##  Max.   :95.0   4/2/2006 0:00:00 :   981   03:00:00 PM:  6700
##                 (Other)          :895866   (Other)    :857229
##    TIME_ZONE          COUNTY           COUNTYNAME         STATE
##  CST    :547493   Min.   :  0.0   JEFFERSON :  7840   TX     : 83728
##  EST    :245558   1st Qu.: 31.0   WASHINGTON:  7603   KS     : 53440
##  MST    : 68390   Median : 75.0   JACKSON   :  6660   OK     : 46802
##  PST    : 28302   Mean   :100.6   FRANKLIN  :  6256   MO     : 35648
##  AST    :  6360   3rd Qu.:131.0   LINCOLN   :  5937   IA     : 31069
##  HST    :  2563   Max.   :873.0   MADISON   :  5632   NE     : 30271
```

```
##  (Other):  3631                   (Other)   :862369   (Other):621339
##                EVTYPE         BGN_RANGE           BGN_AZI
##  HAIL             :288661   Min.   :   0.000          :547332
##  TSTM WIND        :219940   1st Qu.:   0.000   N      : 86752
##  THUNDERSTORM WIND: 82563   Median :   0.000   W      : 38446
##  TORNADO          : 60652   Mean   :   1.484   S      : 37558
##  FLASH FLOOD      : 54277   3rd Qu.:   1.000   E      : 33178
##  FLOOD            : 25326   Max.   :3749.000   NW     : 24041
##  (Other)          :170878                      (Other):134990
##          BGN_LOCATI                  END_DATE             END_TIME
##               :287743                      :243411             :238978
##  COUNTYWIDE   : 19680   4/27/2011 0:00:00:  1214   06:00:00 PM:  9802
##  Countywide   :   993   5/25/2011 0:00:00:  1196   05:00:00 PM:  8314
##  SPRINGFIELD  :   843   6/9/2011 0:00:00 :  1021   04:00:00 PM:  8104
##  SOUTH PORTION:   810   4/4/2011 0:00:00 :  1007   12:00:00 PM:  7483
##  NORTH PORTION:   784   5/30/2004 0:00:00:   998   11:59:00 PM:  7184
##  (Other)      :591444   (Other)          :653450   (Other)    :622432
##    COUNTY_END COUNTYENDN       END_RANGE           END_AZI
##  Min.   :0    Mode:logical   Min.   :  0.0000          :724837
##  1st Qu.:0    NA's:902297    1st Qu.:  0.0000   N      : 28082
##  Median :0                   Median :  0.0000   S      : 22510
##  Mean   :0                   Mean   :  0.9862   W      : 20119
##  3rd Qu.:0                   3rd Qu.:  0.0000   E      : 20047
##  Max.   :0                   Max.   :925.0000   NE     : 14606
##                                                 (Other): 72096
##            END_LOCATI         LENGTH              WIDTH
##                 :499225   Min.   :   0.0000   Min.   :   0.000
##  COUNTYWIDE     : 19731   1st Qu.:   0.0000   1st Qu.:   0.000
##  SOUTH PORTION  :   833   Median :   0.0000   Median :   0.000
##  NORTH PORTION  :   780   Mean   :   0.2301   Mean   :   7.503
##  CENTRAL PORTION:   617   3rd Qu.:   0.0000   3rd Qu.:   0.000
##  SPRINGFIELD    :   575   Max.   :2315.0000   Max.   :4400.000
##  (Other)        :380536
##        F               MAG            FATALITIES          INJURIES
##  Min.   :0.0      Min.   :    0.0   Min.   :  0.0000   Min.   :   0.0000
##  1st Qu.:0.0      1st Qu.:    0.0   1st Qu.:  0.0000   1st Qu.:   0.0000
##  Median :1.0      Median :   50.0   Median :  0.0000   Median :   0.0000
##  Mean   :0.9      Mean   :   46.9   Mean   :  0.0168   Mean   :   0.1557
##  3rd Qu.:1.0      3rd Qu.:   75.0   3rd Qu.:  0.0000   3rd Qu.:   0.0000
##  Max.   :5.0      Max.   :22000.0   Max.   :583.0000   Max.   :1700.0000
##  NA's   :843563
##     PROPDMG          PROPDMGEXP        CROPDMG          CROPDMGEXP
##  Min.   :   0.00          :465934   Min.   :  0.000          :618413
##  1st Qu.:   0.00   K      :424665   1st Qu.:  0.000   K      :281832
##  Median :   0.00   M      : 11330   Median :  0.000   M      :  1994
##  Mean   :  12.06   0      :   216   Mean   :  1.527   k      :    21
##  3rd Qu.:   0.50   B      :    40   3rd Qu.:  0.000   0      :    19
##  Max.   :5000.00   5      :    28   Max.   :990.000   B      :     9
##                    (Other):    84                     (Other):     9
##       WFO                                       STATEOFFIC
##         :142069                                       :248769
##  OUN    : 17393   TEXAS, North                         : 12193
##  JAN    : 13889   ARKANSAS, Central and North Central: 11738
##  LWX    : 13174   IOWA, Central                        : 11345
```

```
##  PHI    : 12551   KANSAS, Southwest                  : 11212
##  TSA    : 12483   GEORGIA, North and Central         : 11120
##  (Other):690738   (Other)                            :595920
##
##
##
##  GREATER RENO / CARSON CITY / M - GREATER RENO / CARSON CITY / M
##  GREATER LAKE TAHOE AREA - GREATER LAKE TAHOE AREA
##  JEFFERSON - JEFFERSON
##  MADISON - MADISON
##  (Other)
##     LATITUDE      LONGITUDE        LATITUDE_E     LONGITUDE_
##  Min.   :   0   Min.   :-14451   Min.   :   0   Min.   :-14455
##  1st Qu.:2802   1st Qu.:  7247   1st Qu.:   0   1st Qu.:     0
##  Median :3540   Median :  8707   Median :   0   Median :     0
##  Mean   :2875   Mean   :  6940   Mean   :1452   Mean   :  3509
##  3rd Qu.:4019   3rd Qu.:  9605   3rd Qu.:3549   3rd Qu.:  8735
##  Max.   :9706   Max.   : 17124   Max.   :9706   Max.   :106220
##  NA's   :47                      NA's   :40
##                                          REMARKS           REFNUM
##                                             :287433   Min.   :     1
##                                             : 24013   1st Qu.:225575
##  Trees down.\n                              :  1110   Median :451149
##  Several trees were blown down.\n           :   568   Mean   :451149
##  Trees were downed.\n                       :   446   3rd Qu.:676723
##  Large trees and power lines were blown down.\n:   432   Max.   :902297
##  (Other)                                    :588295
```

```
names(stormDT)
```

```
##  [1] "STATE__"    "BGN_DATE"   "BGN_TIME"   "TIME_ZONE"  "COUNTY"
##  [6] "COUNTYNAME" "STATE"      "EVTYPE"     "BGN_RANGE"  "BGN_AZI"
## [11] "BGN_LOCATI" "END_DATE"   "END_TIME"   "COUNTY_END" "COUNTYENDN"
## [16] "END_RANGE"  "END_AZI"    "END_LOCATI" "LENGTH"     "WIDTH"
## [21] "F"          "MAG"        "FATALITIES" "INJURIES"   "PROPDMG"
## [26] "PROPDMGEXP" "CROPDMG"    "CROPDMGEXP" "WFO"        "STATEOFFIC"
## [31] "ZONENAMES"  "LATITUDE"   "LONGITUDE"  "LATITUDE_E" "LONGITUDE_"
## [36] "REMARKS"    "REFNUM"
```

Subsetting data

```
cols2Remove <- colnames(stormDT[, !c("EVTYPE"
  , "FATALITIES"
  , "INJURIES"
  , "PROPDMG"
  , "PROPDMGEXP"
  , "CROPDMG"
  , "CROPDMGEXP")])

stormDT[, c(cols2Remove) := NULL]

stormDT <- stormDT[(EVTYPE != "?" &
             (INJURIES > 0 | FATALITIES > 0 | PROPDMG > 0 | CROPDMG > 0)), c("EVTYPE"
                                                                             , "FATALITIES"
```

```
                                                                    , "INJURIES"
                                                                    , "PROPDMG"
                                                                    , "PROPDMGEXP"
                                                                    , "CROPDMG"
                                                                    , "CROPDMGEXP") ]
```

Mapping data includes converting letters (thousand K, million M, billion B) to real numbers

```
cols <- c("PROPDMGEXP", "CROPDMGEXP")
stormDT[,  (cols) := c(lapply(.SD, toupper)), .SDcols = cols]

propDmgKey <-  c("\"\"" = 10^0,
                 "-" = 10^0,
                 "+" = 10^0,
                 "0" = 10^0,
                 "1" = 10^1,
                 "2" = 10^2,
                 "3" = 10^3,
                 "4" = 10^4,
                 "5" = 10^5,
                 "6" = 10^6,
                 "7" = 10^7,
                 "8" = 10^8,
                 "9" = 10^9,
                 "H" = 10^2,
                 "K" = 10^3,
                 "M" = 10^6,
                 "B" = 10^9)

cropDmgKey <-  c("\"\"" = 10^0,
                 "?" = 10^0,
                 "0" = 10^0,
                 "K" = 10^3,
                 "M" = 10^6,
                 "B" = 10^9)

stormDT[, PROPDMGEXP := propDmgKey[as.character(stormDT[,PROPDMGEXP])]]
stormDT[is.na(PROPDMGEXP), PROPDMGEXP := 10^0 ]

stormDT[, CROPDMGEXP := cropDmgKey[as.character(stormDT[,CROPDMGEXP])] ]
stormDT[is.na(CROPDMGEXP), CROPDMGEXP := 10^0 ]
```

Making new column

```
stormDT <- stormDT[, .(EVTYPE, FATALITIES, INJURIES, PROPDMG, PROPDMGEXP, propCost = PROPDMG * PROPDMGE
```

Total property and crop cost

```
totalCostDT <- stormDT[, .(propCost = sum(propCost), cropCost = sum(cropCost), Total_Cost = sum(propCos

totalCostDT <- totalCostDT[order(-Total_Cost), ]

totalCostDT <- totalCostDT[1:10, ]

head(totalCostDT, 10)
```

```
##                EVTYPE     propCost    cropCost   Total_Cost
##  1:             FLOOD 144657709807  5661968450 150319678257
##  2: HURRICANE/TYPHOON  69305840000  2607872800  71913712800
##  3:           TORNADO  56947380677   414953270  57362333947
##  4:       STORM SURGE  43323536000        5000  43323541000
##  5:              HAIL  15735267513  3025954473  18761221986
##  6:       FLASH FLOOD  16822673979  1421317100  18243991079
##  7:           DROUGHT   1046106000 13972566000  15018672000
##  8:         HURRICANE  11868319010  2741910000  14610229010
##  9:       RIVER FLOOD   5118945500  5029459000  10148404500
## 10:         ICE STORM   3944927860  5022113500   8967041360
```

Total fatalities and injuries

```
totalInjuriesDT <- stormDT[, .(FATALITIES = sum(FATALITIES), INJURIES = sum(INJURIES), totals = sum(FAT

totalInjuriesDT <- totalInjuriesDT[order(-FATALITIES), ]

totalInjuriesDT <- totalInjuriesDT[1:10, ]

head(totalInjuriesDT, 10)
```

```
##             EVTYPE FATALITIES INJURIES totals
##  1:        TORNADO       5633    91346  96979
##  2: EXCESSIVE HEAT       1903     6525   8428
##  3:    FLASH FLOOD        978     1777   2755
##  4:           HEAT        937     2100   3037
##  5:      LIGHTNING        816     5230   6046
##  6:      TSTM WIND        504     6957   7461
##  7:          FLOOD        470     6789   7259
##  8:    RIP CURRENT        368      232    600
##  9:      HIGH WIND        248     1137   1385
## 10:      AVALANCHE        224      170    394
```

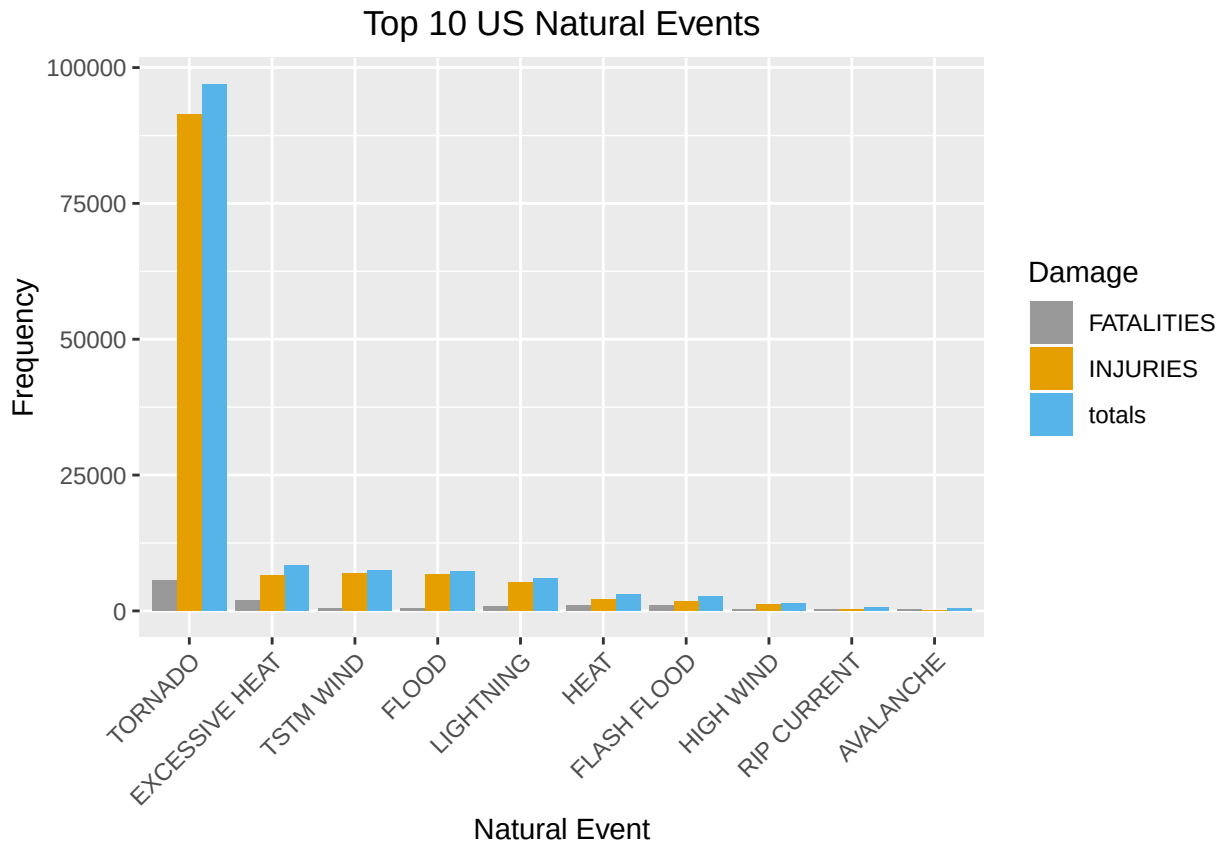
Results

Most harmful events to population health

```
bad_stuff <- melt(totalInjuriesDT, id.vars="EVTYPE", variable.name = "bad_thing")
head(bad_stuff, 10)
```

```
##             EVTYPE  bad_thing value
##  1:        TORNADO FATALITIES  5633
##  2: EXCESSIVE HEAT FATALITIES  1903
##  3:    FLASH FLOOD FATALITIES   978
##  4:           HEAT FATALITIES   937
##  5:      LIGHTNING FATALITIES   816
##  6:      TSTM WIND FATALITIES   504
##  7:          FLOOD FATALITIES   470
##  8:    RIP CURRENT FATALITIES   368
##  9:      HIGH WIND FATALITIES   248
## 10:      AVALANCHE FATALITIES   224
```

```
healthChart <- ggplot(bad_stuff, aes(x=reorder(EVTYPE, -value), y=value))
healthChart = healthChart + geom_bar(stat="identity", aes(fill=bad_thing), position="dodge")
healthChart = healthChart + theme(axis.text.x = element_text(angle=45, hjust=1))
healthChart = healthChart + xlab("Natural Event")
healthChart = healthChart + ylab("Frequency")
healthChart = healthChart + ggtitle("Top 10 US Natural Events") + theme(plot.title = element_text(hjust
healthChart
```



Most damaged events to economic

```
econ_consequences <- melt(totalCostDT, id.vars="EVTYPE", variable.name = "Damage_Type")
head(econ_consequences, 10)
```

```
##                EVTYPE Damage_Type        value
##  1:             FLOOD    propCost 144657709807
##  2: HURRICANE/TYPHOON    propCost  69305840000
##  3:           TORNADO    propCost  56947380677
##  4:       STORM SURGE    propCost  43323536000
##  5:              HAIL    propCost  15735267513
##  6:       FLASH FLOOD    propCost  16822673979
##  7:           DROUGHT    propCost   1046106000
##  8:         HURRICANE    propCost  11868319010
##  9:       RIVER FLOOD    propCost   5118945500
## 10:         ICE STORM    propCost   3944927860
```

```
econChart <- ggplot(econ_consequences, aes(x=reorder(EVTYPE, -value), y=value))
econChart = econChart + theme(axis.text.x = element_text(angle=45, hjust=1))
econChart = econChart + geom_bar(stat="identity", aes(fill=Damage_Type), position="dodge")
econChart = econChart + xlab("Event")
econChart = econChart + ylab("Total Cost")
econChart = econChart + ggtitle("Top 10 US Events Impacting to Economic") + theme(plot.title = element_
econChart
```