Reproducible Research: Peer Assessment 2

RamKamal Tripathi

Tuesday, July 21, 2015

Impact of Severe Weather Events on Public Health and Economy in the United States

Synonpsis

In this report, we aim to analyze the impact of different weather events on public health and economy based on the storm database collected from the U.S. National Oceanic and Atmospheric Administration's (NOAA) from 1950 - 2011. We will use the estimates of fatalities, injuries, property and crop damage to decide which types of event are most harmful to the population health and economy. From these data, we found that excessive heat and tornado are most harmful with respect to population health, while flood, drought, and hurricane/typhoon have the greatest economic consequences.

Basic settings

```
## Loading required package: gridExtra
```

Read the file. If the data already exists in the working environment, we do not need to load it again. Otherwise, we read the csv file.

```r
stormData <- read.csv("repdata-data-StormData.csv", sep = ",")

dim(stormData)
```

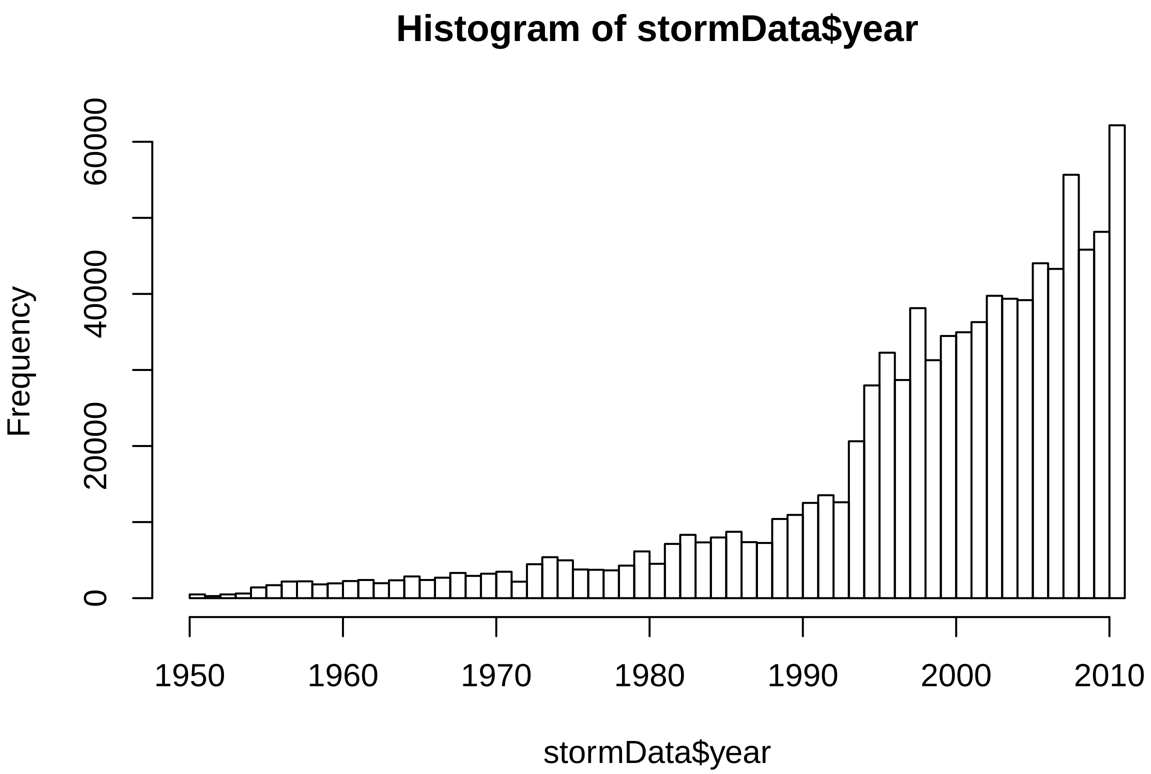
```
## [1] 902297     37
```

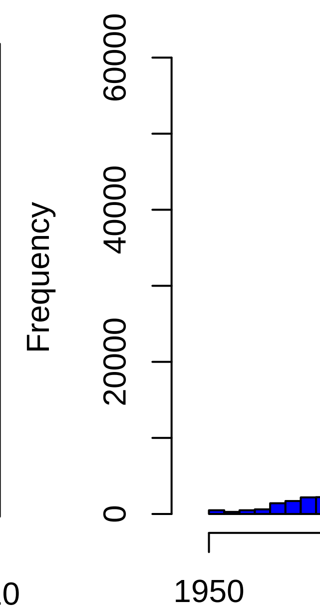
```r
head(stormData, n = 2)
```

```
##   STATE__           BGN_DATE BGN_TIME TIME_ZONE COUNTY COUNTYNAME STATE
## 1       1 4/18/1950 0:00:00     0130       CST     97     MOBILE    AL
## 2       1 4/18/1950 0:00:00     0145       CST      3    BALDWIN    AL
##    EVTYPE BGN_RANGE BGN_AZI BGN_LOCATI END_DATE END_TIME COUNTY_END
## 1 TORNADO         0                                               0
## 2 TORNADO         0                                               0
##   COUNTYENDN END_RANGE END_AZI END_LOCATI LENGTH WIDTH F MAG FATALITIES
## 1         NA         0                        14   100 3   0          0
## 2         NA         0                         2   150 2   0          0
##   INJURIES PROPDMG PROPDMGEXP CROPDMG CROPDMGEXP WFO STATEOFFIC ZONENAMES
## 1       15    25.0          K       0
## 2        0     2.5          K       0
##   LATITUDE LONGITUDE LATITUDE_E LONGITUDE_ REMARKS REFNUM
## 1     3040      8812       3051       8806              1
## 2     3042      8755          0          0              2
```

There are 902297 rows and 37 columns in total. The events in the database start in the year 1950 and end in November 2011. In the earlier years of the database there are generally fewer events recorded, most likely due to a lack of good records. More recent years should be considered more complete.

```
if (dim(stormData)[2] == 37) {
    stormData$year <- as.numeric(format(as.Date(stormData$BGN_DATE, format = "%m/%d/%Y %H:%M:%S"), "%Y"
}
plot(hist(stormData$year, breaks = 50 ) ,xlab="Storm Data per year" , main="Storm Data per year" ,col =
```





Based on the above histogram, we see that the number of events tracked starts to significantly increase around 1995. So, we use the subset of the data from 1990 to 2011 to get most out of good records.

```
storm <- stormData[stormData$year >= 1995, ]
dim(storm)
```

```
## [1] 681500     38
```

Now, there are 681500 rows and 38 columns in total.

Impact on Public Health In this section, we check the number of **fatalities** and **injuries** that are caused by the severe weather events. We would like to get the first 15 most severe types of weather events.

```
sum.fat.Event <- aggregate(FATALITIES~EVTYPE,stormData,sum)
sum.fat.Event.sort <- sum.fat.Event[order(-sum.fat.Event$FATALITIES),]
fatalities <- sum.fat.Event.sort[1:20,]

sum.inj.Event <- aggregate(INJURIES~EVTYPE,stormData,sum)
sum.inj.event.sort <- sum.inj.Event[order(-sum.inj.Event$INJURIES),]
injuries <- sum.inj.event.sort[1:20,]
```

Impact on Economy We will convert the **property damage** and **crop damage** data into comparable numerical forms according to the meaning of units described in the code book (Storm Events). Both PROPDMGEXP and CROPDMGEXP columns record a multiplier for each observation where we have Hundred (H), Thousand (K), Million (M) and Billion (B).

```
convertHelper <- function(dataset = storm, fieldName, newFieldName) {
    totalLen <- dim(dataset)[2]
    index <- which(colnames(dataset) == fieldName)
    dataset[, index] <- as.character(dataset[, index])
    logic <- !is.na(toupper(dataset[, index]))
    dataset[logic & toupper(dataset[, index]) == "B", index] <- "9"
    dataset[logic & toupper(dataset[, index]) == "M", index] <- "6"
    dataset[logic & toupper(dataset[, index]) == "K", index] <- "3"
    dataset[logic & toupper(dataset[, index]) == "H", index] <- "2"
    dataset[logic & toupper(dataset[, index]) == "", index] <- "0"
    dataset[, index] <- as.numeric(dataset[, index])
    dataset[is.na(dataset[, index]), index] <- 0
    dataset <- cbind(dataset, dataset[, index - 1] * 10^dataset[, index])
    names(dataset)[totalLen + 1] <- newFieldName
    return(dataset)
}

storm.Convertion <- convertHelper(storm, "PROPDMGEXP", "propertyDamage")
```

```
## Warning in convertHelper(storm, "PROPDMGEXP", "propertyDamage"): NAs
## introduced by coercion
```

```
storm.Convertion <- convertHelper(storm.Convertion, "CROPDMGEXP", "cropDamage")
```

```
## Warning in convertHelper(storm.Convertion, "CROPDMGEXP", "cropDamage"): NAs
## introduced by coercion
```

```
names(storm.Convertion)
```

```
##  [1] "STATE__"        "BGN_DATE"       "BGN_TIME"       "TIME_ZONE"
##  [5] "COUNTY"         "COUNTYNAME"     "STATE"          "EVTYPE"
##  [9] "BGN_RANGE"      "BGN_AZI"        "BGN_LOCATI"     "END_DATE"
## [13] "END_TIME"       "COUNTY_END"     "COUNTYENDN"     "END_RANGE"
## [17] "END_AZI"        "END_LOCATI"     "LENGTH"         "WIDTH"
## [21] "F"              "MAG"            "FATALITIES"     "INJURIES"
## [25] "PROPDMG"        "PROPDMGEXP"     "CROPDMG"        "CROPDMGEXP"
## [29] "WFO"            "STATEOFFIC"     "ZONENAMES"      "LATITUDE"
## [33] "LONGITUDE"      "LATITUDE_E"     "LONGITUDE_"     "REMARKS"
## [37] "REFNUM"         "year"           "propertyDamage" "cropDamage"
```

```
options(scipen=999) #Disable scientific notation in R


property.damage.sum <- aggregate(propertyDamage~EVTYPE,storm.Convertion,sum)
property.damage.order <-property.damage.sum[order(-property.damage.sum$propertyDamage),]
property <- property.damage.order [1:20,]
```

```r
crop.damage.sum <- aggregate(cropDamage~EVTYPE,storm.Convertion,sum)
crop.damage.order <-crop.damage.sum[order(-crop.damage.sum$cropDamage),]
crop <- crop.damage.order [1:20,]
```

Results

As for the impact on public health, we have got two sorted lists of severe weather events below by the number of people badly affected.

```r
fatalities
```

```
##                      EVTYPE FATALITIES
## 834                 TORNADO       5633
## 130          EXCESSIVE HEAT       1903
## 153             FLASH FLOOD        978
## 275                    HEAT        937
## 464               LIGHTNING        816
## 856               TSTM WIND        504
## 170                   FLOOD        470
## 585             RIP CURRENT        368
## 359               HIGH WIND        248
## 19                AVALANCHE        224
## 972            WINTER STORM        206
## 586            RIP CURRENTS        204
## 278               HEAT WAVE        172
## 140            EXTREME COLD        160
## 760       THUNDERSTORM WIND        133
## 310              HEAVY SNOW        127
## 141 EXTREME COLD/WIND CHILL        125
## 676             STRONG WIND        103
## 30                 BLIZZARD        101
## 350               HIGH SURF        101
```

```r
injuries
```

```
##                 EVTYPE INJURIES
## 834            TORNADO    91346
## 856          TSTM WIND     6957
## 170              FLOOD     6789
## 130     EXCESSIVE HEAT     6525
## 464          LIGHTNING     5230
## 275               HEAT     2100
## 427          ICE STORM     1975
## 153        FLASH FLOOD     1777
## 760  THUNDERSTORM WIND     1488
## 244               HAIL     1361
## 972       WINTER STORM     1321
## 411  HURRICANE/TYPHOON     1275
## 359          HIGH WIND     1137
## 310         HEAVY SNOW     1021
## 957           WILDFIRE      911
## 786 THUNDERSTORM WINDS      908
```

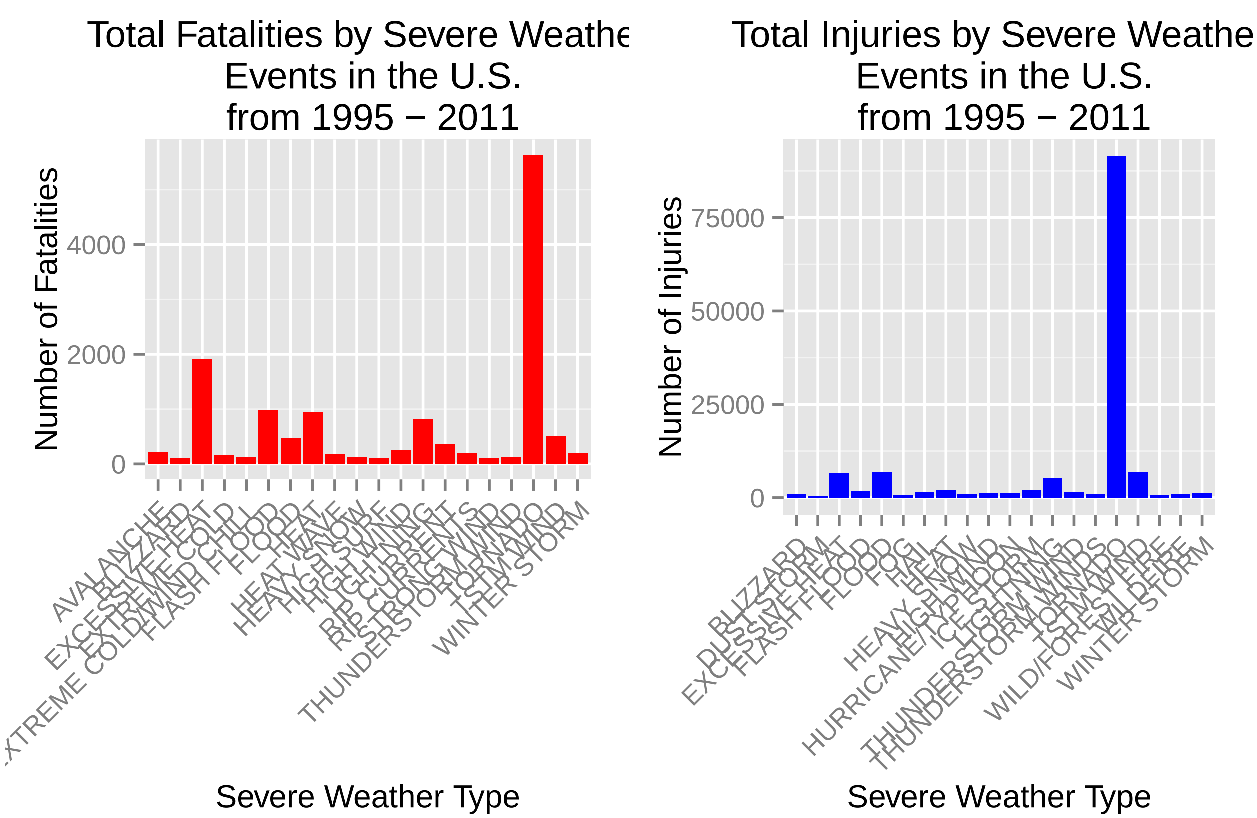
```
## 30          BLIZZARD       805
## 188              FOG       734
## 955  WILD/FOREST FIRE      545
## 117       DUST STORM       440
```

And the following is a pair of graphs of total fatalities and total injuries affected by these severe weather events.

```r
fatalitiesPlot <- qplot(EVTYPE, data = fatalities, weight = FATALITIES, geom = "bar", binwidth = 1 ) +
    geom_bar(fill="red" ) +
    scale_y_continuous("Number of Fatalities") +
    theme(axis.text.x = element_text(angle = 45,
    hjust = 1)) + xlab("Severe Weather Type") +
    ggtitle("Total Fatalities by Severe Weather\n Events in the U.S.\n from 1995 - 2011")


injuriesPlot <- qplot(EVTYPE, data = injuries, weight = INJURIES, geom = "bar", binwidth = 1) +
    geom_bar(fill="blue" ) +
    scale_y_continuous("Number of Injuries") +
    theme(axis.text.x = element_text(angle = 45,
    hjust = 1)) + xlab("Severe Weather Type") +
    ggtitle("Total Injuries by Severe Weather\n Events in the U.S.\n from 1995 - 2011")

grid.arrange(fatalitiesPlot, injuriesPlot, ncol = 2)
```



Based on the above histograms, we find that **excessive heat** and **tornado** cause most fatalities; **tornato** causes most injuries in the United States from 1995 to 2011.

As for the impact on economy, we have got two sorted lists below by the amount of money cost by damages.

```
property
```

```
##                         EVTYPE propertyDamage
## 144                      FLOOD   144022037057
## 313          HURRICANE/TYPHOON    69305840000
## 519                STORM SURGE    43193536000
## 666                    TORNADO    24935939545
## 134                FLASH FLOOD    16047794571
## 206                       HAIL    15048722103
## 306                  HURRICANE    11812819010
## 677             TROPICAL STORM     7653335550
## 288                  HIGH WIND     5259785375
## 773                   WILDFIRE     4759064000
## 520           STORM SURGE/TIDE     4641188000
## 683                  TSTM WIND     4482361440
## 326                  ICE STORM     3643555810
## 607          THUNDERSTORM WIND     3399282992
## 310             HURRICANE OPAL     3172846000
## 771           WILD/FOREST FIRE     3001812500
## 247 HEAVY RAIN/SEVERE WEATHER     2500000000
## 787               WINTER STORM     1538047250
## 479        SEVERE THUNDERSTORM     1200310000
## 84                     DROUGHT     1046106000
```

```
crop
```

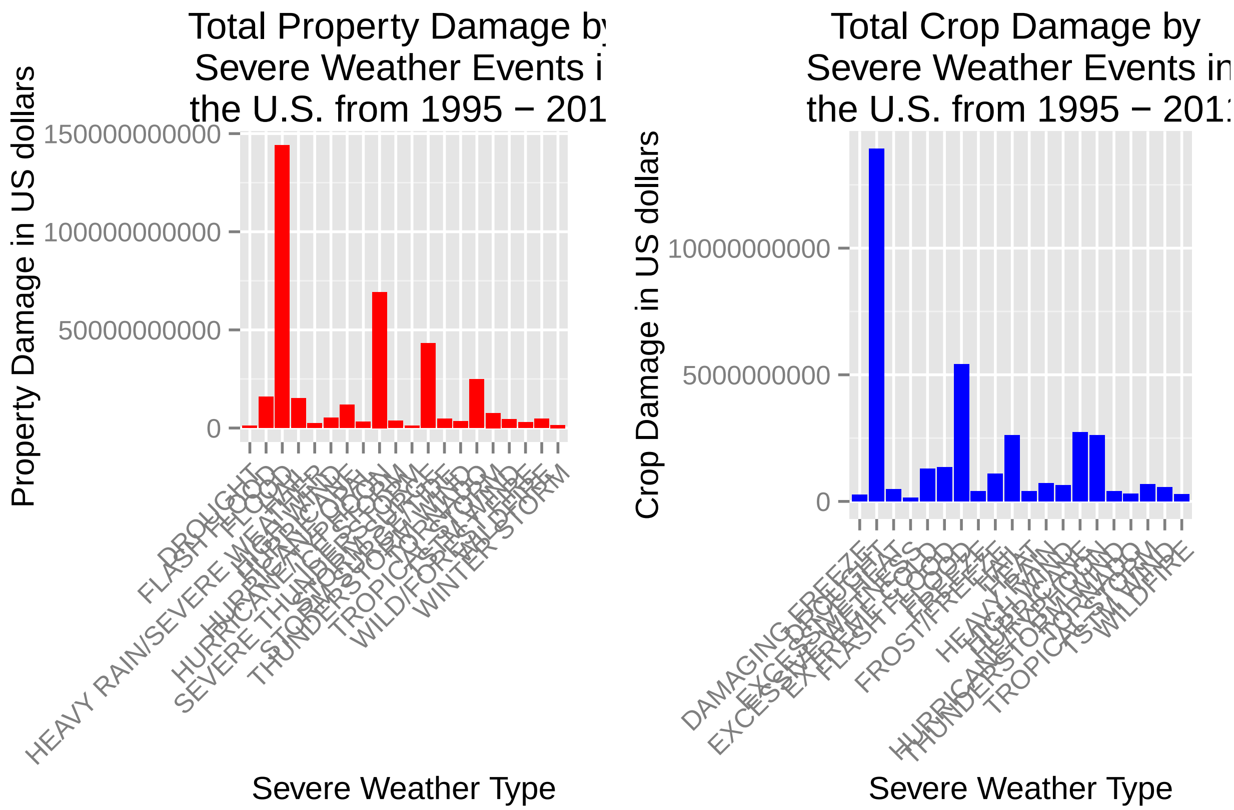
```
##                EVTYPE  cropDamage
## 84            DROUGHT 13922066000
## 144             FLOOD  5422810400
## 306         HURRICANE  2741410000
## 206              HAIL  2614127070
## 313 HURRICANE/TYPHOON  2607872800
## 134       FLASH FLOOD  1343915000
## 121      EXTREME COLD  1292473000
## 179      FROST/FREEZE  1094086000
## 241        HEAVY RAIN   728399800
## 677    TROPICAL STORM   677836000
## 288         HIGH WIND   633561300
## 683         TSTM WIND   553947350
## 112    EXCESSIVE HEAT   492402000
## 607 THUNDERSTORM WIND   414354000
## 231              HEAT   401411500
## 159            FREEZE   396225000
## 666           TORNADO   296595770
## 773          WILDFIRE   295472800
## 76    DAMAGING FREEZE   262100000
## 117 EXCESSIVE WETNESS   142000000
```

And the following is a pair of graphs of total property damage and total crop damage affected by these severe weather events.

```
propertyPlot <- qplot(EVTYPE, data = property, weight = propertyDamage, geom = "bar", binwidth = 1) +
    theme(axis.text.x = element_text(angle = 45, hjust = 1)) + scale_y_continuous("Property Damage in US
    xlab("Severe Weather Type") + ggtitle("Total Property Damage by\n Severe Weather Events in\n the U.S

cropPlot<- qplot(EVTYPE, data = crop, weight = cropDamage, geom = "bar", binwidth = 1) +
    geom_bar(fill="blue" ) +
    theme(axis.text.x = element_text(angle = 45, hjust = 1)) + scale_y_continuous("Crop Damage in US do
    xlab("Severe Weather Type") + ggtitle("Total Crop Damage by \nSevere Weather Events in\n the U.S. f

grid.arrange(propertyPlot, cropPlot, ncol = 2)
```



Based on the above histograms, we find that **flood** and **hurricane/typhoon** cause most property damage; **drought** and **flood** causes most crop damage in the United States from 1995 to 2011.

Conclusion

From these data, we found that **excessive heat** and **tornado** are most harmful with respect to population health, while **flood**, **drought**, and **hurricane/typhoon** have the greatest economic consequences.