Reproducible Research Course Project 2

Charles Yoo

June 17, 2016

Project Title

U.S. National Oceanic and Atmospheric Administration's (NOAA) Storm Database Analysis

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.

Questions to Answer

- Across the United States, which types of events (as indicated in the `EVTYPE` variable) are most harmful with respect to population health?
- Across the United States, which types of events have the greatest economic consequences?

Data Processing

[https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2 (https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2)] (Storm Data) [https://d396qusza40orc.cloudfront.net/repdata%2Fpeer2_doc%2Fpd01016005curr.pdf (https://d396qusza40orc.cloudfront.net/repdata%2Fpeer2_doc%2Fpd01016005curr.pdf)] (National Weather Service Storm Data Documentation) [https://d396qusza40orc.cloudfront.net/repdata%2Fpeer2_doc%2FNCDC%20Storm%20Events-FAQ%20Page.pdf (https://d396qusza40orc.cloudfront.net/repdata%2Fpeer2_doc%2FNCDC%20Storm%20Events-FAQ%20Page.pdf)] (National Climatic Data Center Storm Events FAQ)

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.

Download, extract and read CSV file

```
download.file("https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2",

              dest="stormdata.bz2",
              method="curl")

data <- read.csv(bzfile("stormdata.bz2"),
                header = TRUE,
                sep = ",",
                stringsAsFactors = FALSE)
```

Examine data

```
str(data)
```

```
## 'data.frame':    902297 obs. of  37 variables:
##  $ STATE__   : num  1 1 1 1 1 1 1 1 1 1 ...
##  $ BGN_DATE  : chr  "4/18/1950 0:00:00" "4/18/1950 0:00:00" "2/20/1951 0:00:00" "6/8/
1951 0:00:00" ...
##  $ BGN_TIME  : chr  "0130" "0145" "1600" "0900" ...
##  $ TIME_ZONE : chr  "CST" "CST" "CST" "CST" ...
##  $ COUNTY    : num  97 3 57 89 43 77 9 123 125 57 ...
##  $ COUNTYNAME: chr  "MOBILE" "BALDWIN" "FAYETTE" "MADISON" ...
##  $ STATE     : chr  "AL" "AL" "AL" "AL" ...
##  $ EVTYPE    : chr  "TORNADO" "TORNADO" "TORNADO" "TORNADO" ...
##  $ BGN_RANGE : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ BGN_AZI   : chr  "" "" "" "" ...
##  $ BGN_LOCATI: chr  "" "" "" "" ...
##  $ END_DATE  : chr  "" "" "" "" ...
##  $ END_TIME  : chr  "" "" "" "" ...
##  $ COUNTY_END: num  0 0 0 0 0 0 0 0 0 0 ...
##  $ COUNTYENDN: logi  NA NA NA NA NA NA ...
##  $ END_RANGE : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ END_AZI   : chr  "" "" "" "" ...
##  $ END_LOCATI: chr  "" "" "" "" ...
##  $ LENGTH    : num  14 2 0.1 0 0 1.5 1.5 0 3.3 2.3 ...
##  $ WIDTH     : num  100 150 123 100 150 177 33 33 100 100 ...
##  $ F         : int  3 2 2 2 2 2 2 1 3 3 ...
##  $ MAG       : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ FATALITIES: num  0 0 0 0 0 0 0 0 1 0 ...
##  $ INJURIES  : num  15 0 2 2 2 6 1 0 14 0 ...
##  $ PROPDMG   : num  25 2.5 25 2.5 2.5 2.5 2.5 2.5 25 25 ...
##  $ PROPDMGEXP: chr  "K" "K" "K" "K" ...
##  $ CROPDMG   : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ CROPDMGEXP: chr  "" "" "" "" ...
##  $ WFO       : chr  "" "" "" "" ...
##  $ STATEOFFIC: chr  "" "" "" "" ...
##  $ ZONENAMES : chr  "" "" "" "" ...
##  $ LATITUDE  : num  3040 3042 3340 3458 3412 ...
##  $ LONGITUDE : num  8812 8755 8742 8626 8642 ...
##  $ LATITUDE_E: num  3051 0 0 0 0 ...
##  $ LONGITUDE_: num  8806 0 0 0 0 ...
##  $ REMARKS   : chr  "" "" "" "" ...
##  $ REFNUM    : num  1 2 3 4 5 6 7 8 9 10 ...
```

```
summary(data)
```

```
##     STATE__       BGN_DATE           BGN_TIME          TIME_ZONE
##  Min.   : 1.0   Length:902297      Length:902297      Length:902297
##  1st Qu.:19.0   Class :character   Class :character   Class :character
##  Median :30.0   Mode  :character   Mode  :character   Mode  :character
##  Mean   :31.2
##  3rd Qu.:45.0
##  Max.   :95.0
##
##      COUNTY       COUNTYNAME           STATE              EVTYPE
##  Min.   :  0.0   Length:902297      Length:902297      Length:902297
##  1st Qu.: 31.0   Class :character   Class :character   Class :character
##  Median : 75.0   Mode  :character   Mode  :character   Mode  :character
##  Mean   :100.6
##  3rd Qu.:131.0
##  Max.   :873.0
##
##    BGN_RANGE          BGN_AZI           BGN_LOCATI
##  Min.   :   0.000   Length:902297      Length:902297
##  1st Qu.:   0.000   Class :character   Class :character
##  Median :   0.000   Mode  :character   Mode  :character
##  Mean   :   1.484
##  3rd Qu.:   1.000
##  Max.   :3749.000
##
##    END_DATE           END_TIME           COUNTY_END COUNTYENDN
##  Length:902297      Length:902297      Min.   :0    Mode:logical
##  Class :character   Class :character   1st Qu.:0    NA's:902297
##  Mode  :character   Mode  :character   Median :0
##                                        Mean   :0
##                                        3rd Qu.:0
##                                        Max.   :0
##
##    END_RANGE          END_AZI           END_LOCATI
##  Min.   :  0.0000   Length:902297      Length:902297
##  1st Qu.:  0.0000   Class :character   Class :character
##  Median :  0.0000   Mode  :character   Mode  :character
##  Mean   :  0.9862
##  3rd Qu.:  0.0000
##  Max.   :925.0000
##
##      LENGTH              WIDTH                 F               MAG
##  Min.   :   0.0000   Min.   :   0.000   Min.   :0.0      Min.   :    0.0
##  1st Qu.:   0.0000   1st Qu.:   0.000   1st Qu.:0.0      1st Qu.:    0.0
##  Median :   0.0000   Median :   0.000   Median :1.0      Median :   50.0
##  Mean   :   0.2301   Mean   :   7.503   Mean   :0.9      Mean   :   46.9
##  3rd Qu.:   0.0000   3rd Qu.:   0.000   3rd Qu.:1.0      3rd Qu.:   75.0
##  Max.   :2315.0000   Max.   :4400.000   Max.   :5.0      Max.   :22000.0
##                                         NA's   :843563
##    FATALITIES          INJURIES            PROPDMG
##  Min.   :  0.0000   Min.   :   0.0000   Min.   :   0.00
##  1st Qu.:  0.0000   1st Qu.:   0.0000   1st Qu.:   0.00
##  Median :  0.0000   Median :   0.0000   Median :   0.00
##  Mean   :  0.0168   Mean   :   0.1557   Mean   :  12.06
##  3rd Qu.:  0.0000   3rd Qu.:   0.0000   3rd Qu.:   0.50  
##  Max.   :583.0000   Max.   :1700.0000   Max.   :5000.00  
##                                                          
##   PROPDMGEXP           CROPDMG         CROPDMGEXP       
##  Length:902297      Min.   :  0.000   Length:902297     
##  Class :character   1st Qu.:  0.000   Class :character  
##  Mode  :character   Median :  0.000   Mode  :character  
##                     Mean   :  1.527                     
##                     3rd Qu.:  0.000                     
##                     Max.   :990.000                     
##                                                         
##      WFO             STATEOFFIC         ZONENAMES            LATITUDE   
##  Length:902297      Length:902297      Length:902297      Min.   :   0  
##  Class :character   Class :character   Class :character   1st Qu.:2802  
##  Mode  :character   Mode  :character   Mode  :character   Median :3540  
##                                                           Mean   :2875  
##                                                           3rd Qu.:4019  
##                                                           Max.   :9706  
##                                                           NA's   :47    
##    LONGITUDE        LATITUDE_E     LONGITUDE_       REMARKS         
##  Min.   :-14451   Min.   :   0   Min.   :-14455   Length:902297     
##  1st Qu.:  7247   1st Qu.:   0   1st Qu.:     0   Class :character  
##  Median :  8707   Median :   0   Median :     0   Mode  :character  
##  Mean   :  6940   Mean   :1452   Mean   :  3509                     
##  3rd Qu.:  9605   3rd Qu.:3549   3rd Qu.:  8735                     
##  Max.   : 17124   Max.   :9706   Max.   :106220                     
##                   NA's   :40                                        
##      REFNUM      
##  Min.   :     1  
##  1st Qu.:225575  
##  Median :451149  
##  Mean   :451149  
##  3rd Qu.:676723  
##  Max.   :902297  
## 
```

View data

```
#View(data)
```

Lowercase variable names

```
colnames(data) <- tolower(colnames(data))
```

Change to factor for event types

```
data$evtype <- as.factor(data$evtype)
```

Change dates to POSIXct

```
data$bgn_date <- as.Date(data$bgn_date, "%m/%d/%Y")
data$end_date <- as.Date(data$end_date, "%m/%d/%Y")
```

Subset data to records that have fatalities, injuries related only

```
fData <- subset(data, subset = data$fatalities > 0)
iData <- subset(data, subset = data$injuries > 0)
phyData <- rbind(fData, iData)
```

Subset data to records that have property, crop damage related only

```
pData <- subset(data, subset = data$propdmg > 0)
cData <- subset(data, subset = data$cropdmg > 0)
ecoData <- rbind(pData, cData)
```

Subset data to records that have fatalities, injuries, property and crop damage

```
data <- rbind(phyData, ecoData)
```

View data

```
#View(data)
```

Upper property and crop exponent multipliers Map multipliers to numeric values

```
data$propdmgexp <- toupper(data$propdmgexp)
data$cropdmgexp <- toupper(data$cropdmgexp)

pDmgExp <- 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)

data$propdmgexp <- pDmgExp[as.character(data$propdmgexp)]
data$propdmgexp[is.na(data$propdmgexp)] <- 10^0

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

data$cropdmgexp <- cDmgExp[as.character(data$cropdmgexp)]
data$cropdmgexp[is.na(data$cropdmgexp)] <- 10^0
```

Aggregate data for physical data, calculating totals

```
aggregatedPhysicalData <- aggregate(cbind(fatalities, injuries) ~ evtype, data = data, F
UN = sum)
aggregatedPhysicalData$total <- aggregatedPhysicalData$fatalities + aggregatedPhysicalDa
ta$injuries
```

Determine events with the highest physical impact, by aggregating data

```
# aggregated data greater than 0
aggregatedPhysicalData <- aggregatedPhysicalData[aggregatedPhysicalData$total > 0, ]
# sort in descending order
aggregatedPhysicalData <- aggregatedPhysicalData[order(aggregatedPhysicalData$total, decreasing = TRUE), ]

# renumber rows in descending order
rownames(aggregatedPhysicalData) <- 1:nrow(aggregatedPhysicalData)

# top 10
aggregatedPhysicalDataTop10 <- aggregatedPhysicalData[1:10, ]
```

Calculate property and crop loss based on exponent multipliers

```
data$propertyLoss <- data$propdmg * data$propdmgexp
data$cropLoss <- data$cropdmg * data$cropdmgexp
```

Aggregate data for economic data, calculating totals

```
aggregatedEconomicData <- aggregate(cbind(propertyLoss, cropLoss) ~ evtype, data = data, FUN = sum)
# create total field
aggregatedEconomicData$total <- aggregatedEconomicData$propertyLoss + aggregatedEconomicData$cropLoss
```

Determine events with the highest economic impact

```
# aggregated data greater than 0
aggregatedEconomicData <- aggregatedEconomicData[aggregatedEconomicData$total > 0, ]
# sort in descending order
aggregatedEconomicData <- aggregatedEconomicData[order(aggregatedEconomicData$total, decreasing = TRUE), ]

# renumber rows in descending order
rownames(aggregatedEconomicData) <- 1:nrow(aggregatedEconomicData)

# top 10
aggregatedEconomicDataTop10 <- aggregatedEconomicData[1:10, ]
```

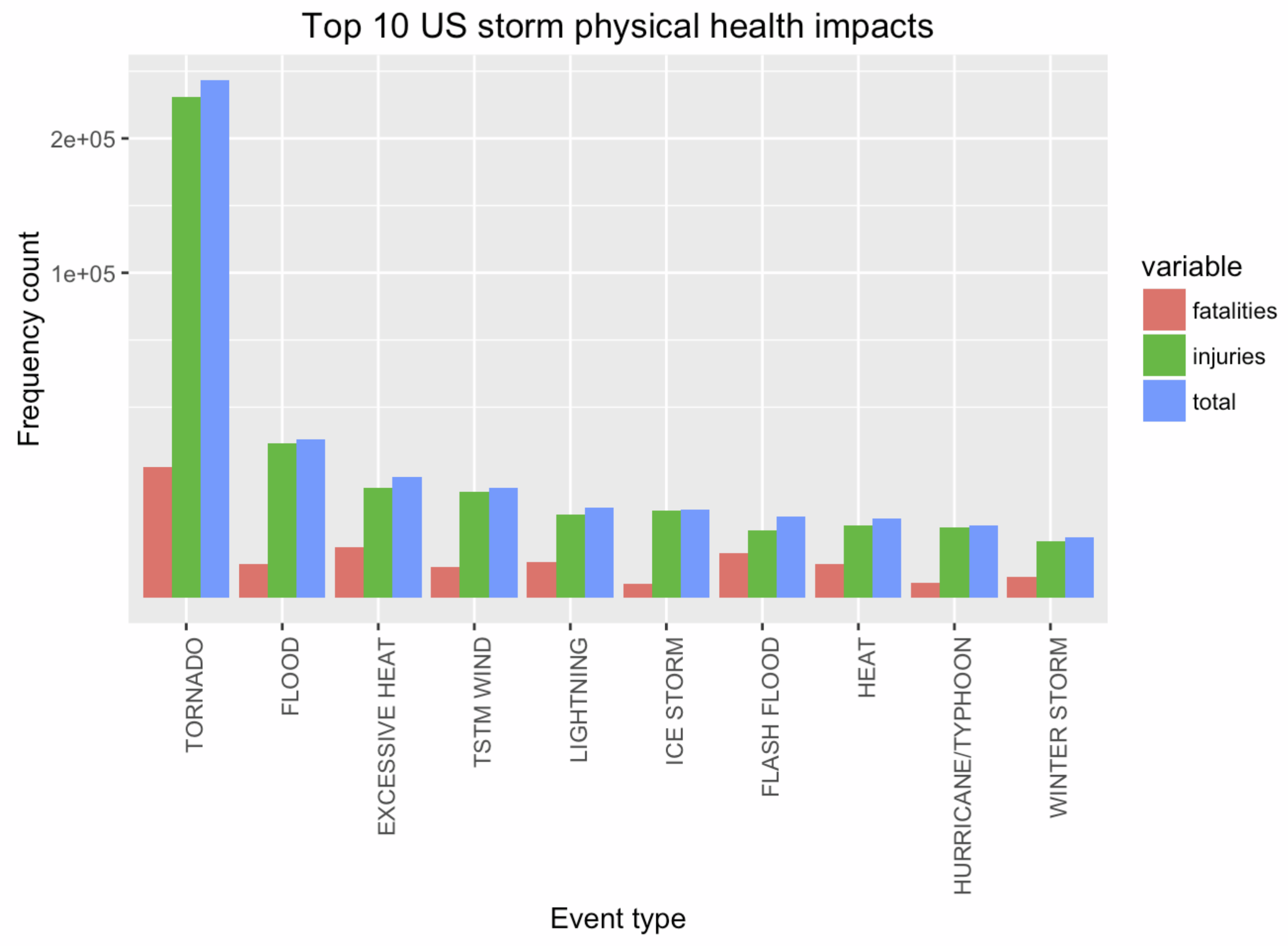
Results

Plot of the top ten event types with the highest fatality and injury counts

```
# melt id variables evtype
physicalDataTop10Melt <- melt(aggregatedPhysicalDataTop10, id.vars = "evtype")

# build ggplot of top 10
physicalChart <- ggplot(physicalDataTop10Melt, aes(x = reorder(evtype, -value), y = valu
e)) +
                        geom_bar(stat = "identity", aes(fill = variable), position = "do
dge") +
                        scale_y_sqrt("Frequency count") +
                        xlab("Event type") +
                        theme(axis.text.x = element_text(angle = 90, hjust = 1)) +
                        ggtitle("Top 10 US storm physical health impacts")

print(physicalChart)
```

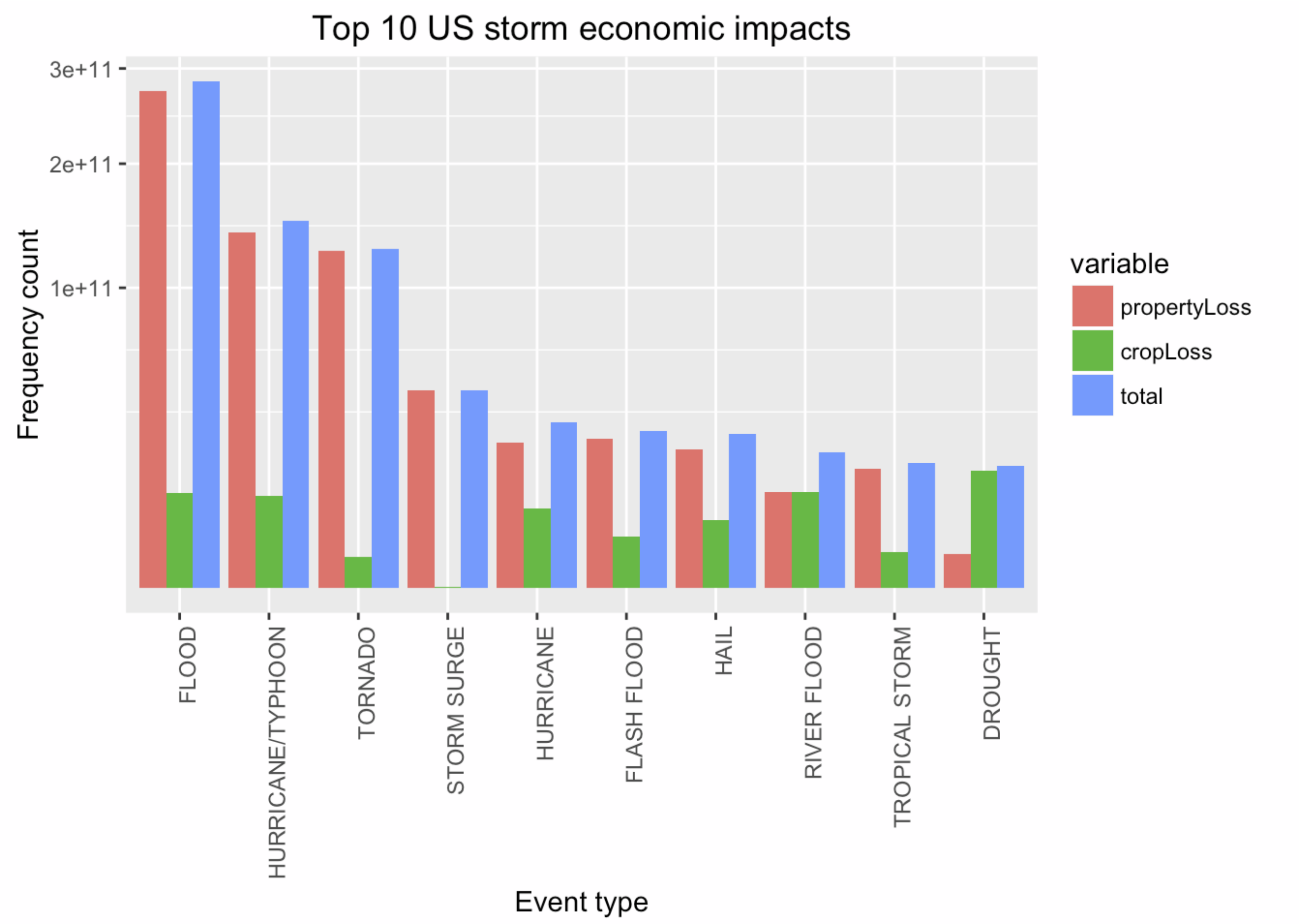


Plot of the top ten event types with the highest property and crop counts

```
# melt id variables evtype
economicDataTop10Melt <- melt(aggregatedEconomicDataTop10, id.vars = "evtype")

# build ggplot of top 10
economicChart <- ggplot(economicDataTop10Melt, aes(x = reorder(evtype, -value), y = valu
e)) +
                        geom_bar(stat = "identity", aes(fill = variable), position = "do
dge") +
                        scale_y_sqrt("Frequency count") +
                        xlab("Event type") +
                        theme(axis.text.x = element_text(angle = 90, hjust = 1)) +
                        ggtitle("Top 10 US storm economic impacts")

print(economicChart)
```



Answers

Across the United States, which types of events (as indicated in the `EVTYPE` variable) are most harmful with respect to population health?

```
Based on the results, tornados are the most harmful with to population health, fatalitie
s and injuries.
```

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

```
Based on the results, floods have the greatest economic consequences.
```

References

```
Reference:
http://rpubs.com/cneiderer/rrCourseProject2
https://rpubs.com/withgemini/25349
https://github.com/paul-reiners/reproducible-research-project-2/blob/master/StormReport.
Rmd
http://54.225.166.221/Drarreg/rrcourseproject2
```