

REPRODUCIBLE RESEARCH COURSE PROJECT 2

2023-05-05

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

Two questions will be answered

Across the United States, which types of events are most harmful with respect to population health? Across the United States, which types of events have the greatest economic consequences?

BASIC SETTINGS

```
#install.packages("ggplot2")
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.2.3
```

```
#install.packages("data.table")
library("data.table")
```

```
## Warning: package 'data.table' was built under R version 4.2.3
```

```
#install.packages("dplyr")
library("dplyr")
```

```
## Warning: package 'dplyr' was built under R version 4.2.3
```

```
##
```

```
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:data.table':
```

```
##
```

```
##      between, first, last
```

```
## The following objects are masked from 'package:stats':  
##  
##   filter, lag
```

```
## The following objects are masked from 'package:base':  
##  
##   intersect, setdiff, setequal, union
```

```
#install.packages("plyr")  
library(plyr)
```

```
## Warning: package 'plyr' was built under R version 4.2.3
```

```
## -----
```

```
## You have loaded plyr after dplyr - this is likely to cause problems.  
## If you need functions from both plyr and dplyr, please load plyr first, then dplyr:  
## library(plyr); library(dplyr)
```

```
## -----
```

```
##  
## Attaching package: 'plyr'
```

```
## The following objects are masked from 'package:dplyr':  
##  
##   arrange, count, desc, failwith, id, mutate, rename, summarise,  
##   summarize
```

```
#install.packages("lubridate")  
library("lubridate")
```

```
## Warning: package 'lubridate' was built under R version 4.2.3
```

```
##  
## Attaching package: 'lubridate'
```

```
## The following objects are masked from 'package:data.table':  
##  
##   hour, isoweek, mday, minute, month, quarter, second, wday, week,  
##   yday, year
```

```
## The following objects are masked from 'package:base':  
##  
##   date, intersect, setdiff, union
```

```
library(R.utils)
```

```
## Warning: package 'R.utils' was built under R version 4.2.3
```

```
## Loading required package: R.oo

## Loading required package: R.methodsS3

## R.methodsS3 v1.8.2 (2022-06-13 22:00:14 UTC) successfully loaded. See ?R.methodsS3 for help.

## R.oo v1.25.0 (2022-06-12 02:20:02 UTC) successfully loaded. See ?R.oo for help.

##
## Attaching package: 'R.oo'

## The following object is masked from 'package:R.methodsS3':
##
##      throw

## The following objects are masked from 'package:methods':
##
##      getClasses, getMethods

## The following objects are masked from 'package:base':
##
##      attach, detach, load, save

## R.utils v2.12.2 (2022-11-11 22:00:03 UTC) successfully loaded. See ?R.utils for help.

##
## Attaching package: 'R.utils'

## The following object is masked from 'package:utils':
##
##      timestamp

## The following objects are masked from 'package:base':
##
##      cat, commandArgs, getOption, isOpen, nullfile, parse, warnings
```

DATA PROCESSING

```
stormDraw <- data.table::fread(input = "repdata_data_StormData.csv.bz2")
str(stormDraw)
```

```
## Classes 'data.table' and '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" ...
## $ CROPDGMG    : num 0 0 0 0 0 0 0 0 0 0 ...
## $ CROPDGMGEXP : 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 ...
## - attr(*, ".internal.selfref")=<externalptr>
```

```
seltData <- c("EVTYPE", "FATALITIES", "INJURIES", "PROPDMG", "PROPDMGEXP", "CROPDMG", "CROPDMGEXP")
stormDTsubset <- stormDraw[,seltData, with = FALSE]
names(stormDTsubset)
```

```
## [1] "EVTYPE"      "FATALITIES" "INJURIES"   "PROPDMG"    "PROPDMGEXP"
## [6] "CROPDMG"     "CROPDMGEXP"
```

```
str(stormDTsubset)
```

```
## Classes 'data.table' and 'data.frame': 902297 obs. of 7 variables:
## $ EVTYPE      : chr "TORNADO" "TORNADO" "TORNADO" "TORNADO" ...
## $ 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 "" "" "" "" ...
## - attr(*, ".internal.selfref")=<externalptr>
```

```
sum(is.na(stormDTsubset$EVTYPE))
```

```
## [1] 0
```

```
sum(is.na(stormDTsubset$FATALITIES))
```

```
## [1] 0
```

```
sum(is.na(stormDTsubset$INJURIES))
```

```
## [1] 0
```

```
sum(is.na(stormDTsubset$PROPDMG))
```

```
## [1] 0
```

```
sum(is.na(stormDTsubset$PROPDMGEXP))
```

```
## [1] 0
```

```
sum(is.na(stormDTsubset$CROPDMG))
```

```
## [1] 0
```

```
sum(is.na(stormDTsubset$CROPDMGEXP))
```

```
## [1] 0
```

```
unique(stormDTsubset$PROPDMGEXP)
```

```
## [1] "K" "M" "" "B" "m" "+" "0" "5" "6" "?" "4" "2" "3" "h" "7" "H" "-" "1" "8"
```

```
stormDTsubset$PROPDMGEXP <- mapvalues(stormDTsubset$PROPDMGEXP, from = c("K", "M", "", "B", "m", "+", "0", "5", "6", "?", "4", "2", "3", "h", "7", "H", "-", "1", "8"), to = c("K", "M", "0", "B", "m", "+", "0", "5", "6", "?", "4", "2", "3", "h", "7", "H", "-", "1", "8"))
stormDTsubset$PROPDMGEXP <- as.numeric(as.character(stormDTsubset$PROPDMGEXP))
stormDTsubset$PROPDMGTOTAL <- (stormDTsubset$PROPDMG * stormDTsubset$PROPDMGEXP)/1000000000
unique(stormDTsubset$CROPDMGEXP)
```

```
## [1] "" "M" "K" "m" "B" "?" "0" "k" "2"
```

```
stormDTsubset$CROPDMGEXP <- mapvalues(stormDTsubset$CROPDMGEXP, from = c("", "M", "K", "m", "B", "?", "0", "k", "2"), to = c("", "M", "K", "m", "B", "?", "0", "k", "2"))
stormDTsubset$CROPDMGEXP <- as.numeric(as.character(stormDTsubset$CROPDMGEXP))
stormDTsubset$CROPDMGTOTAL <- (stormDTsubset$CROPDMG * stormDTsubset$CROPDMGEXP)/1000000000
```

RESULTS

Events for public health variables Fatalities

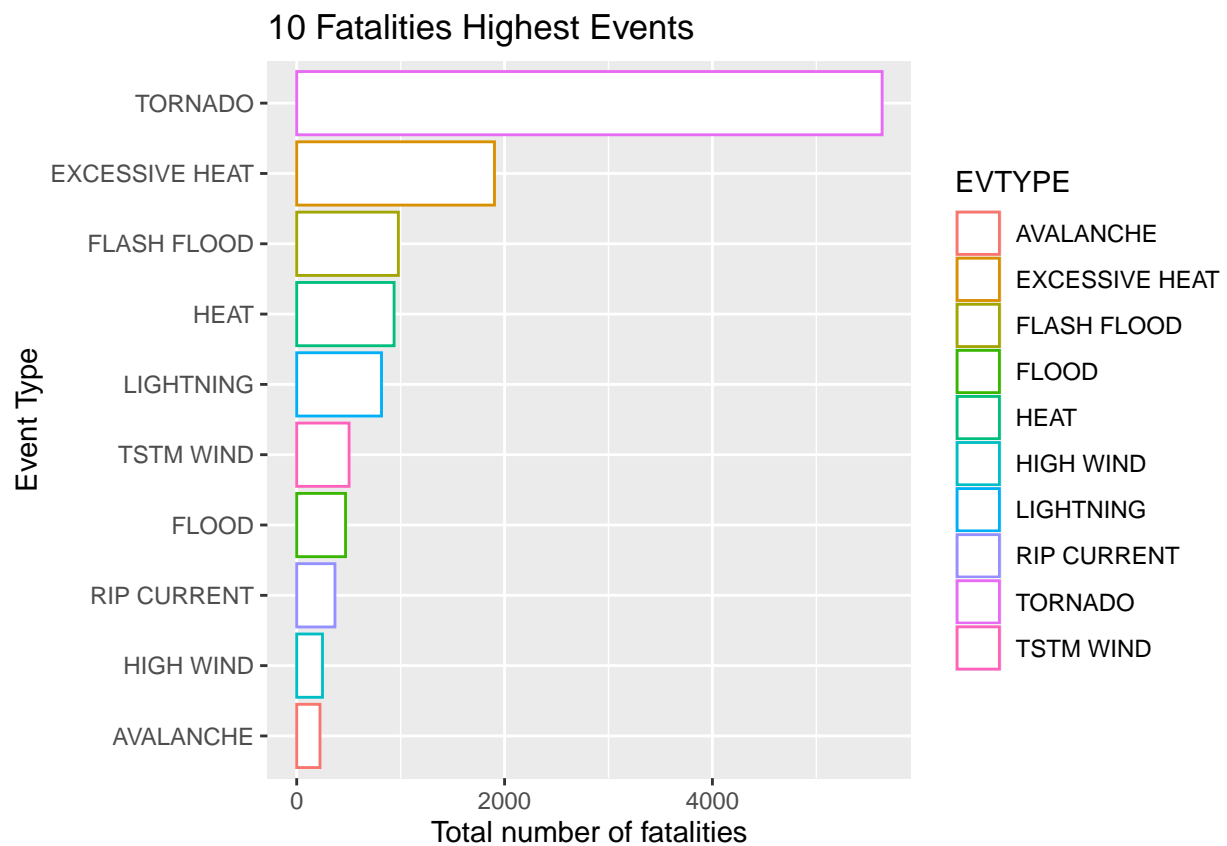
```
aggFatalites <- aggregate(FATALITIES ~ EVTYPE, data = stormDTsubset, FUN="sum")
dim(aggFatalites)
```

```
## [1] 985 2
```

```
evFatalities <- aggFatalites[order(-aggFatalites$FATALITIES), ][1:10, ]
evFatalities
```

```
##           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
```

```
fatalitiesPlot <- ggplot(data=evFatalities, aes(x=reorder(EVTYPE, FATALITIES), y=FATALITIES, color=EVTYPE))
fatalitiesPlot + coord_flip()
```



INJURIES

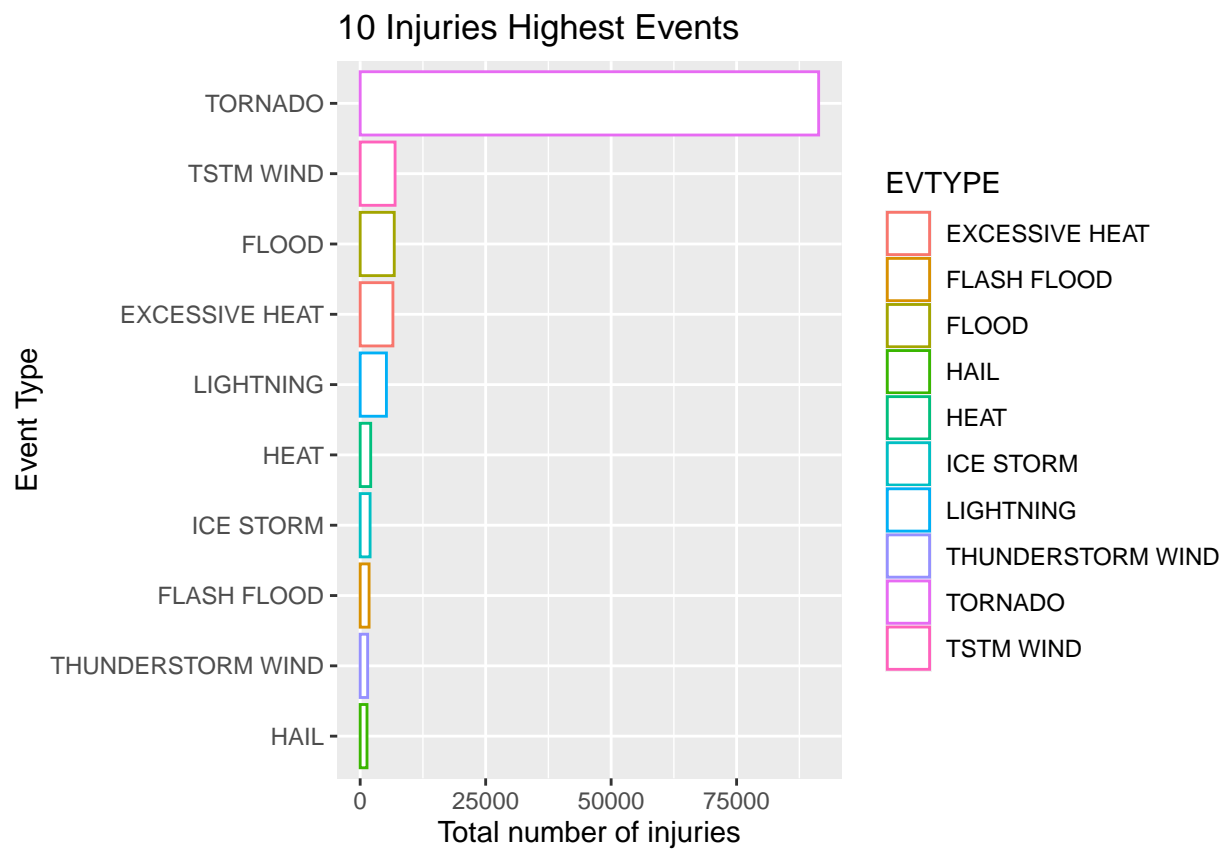
```
aggInjuries <- aggregate(INJURIES ~ EVTYPE, data = stormDTsubset, FUN="sum")
dim(aggInjuries)
```

```
## [1] 985 2
```

```
evInjuries <- aggInjuries[order(-aggInjuries$INJURIES), ][1:10, ]
evInjuries
```

```
##           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
```

```
injuriesPlot <- ggplot(data=evInjuries, aes(x=reorder(EVTYPE, INJURIES), y=INJURIES, color=EVTYPE)) + g
injuriesPlot + coord_flip()
```



PROPERTY DAMAGE

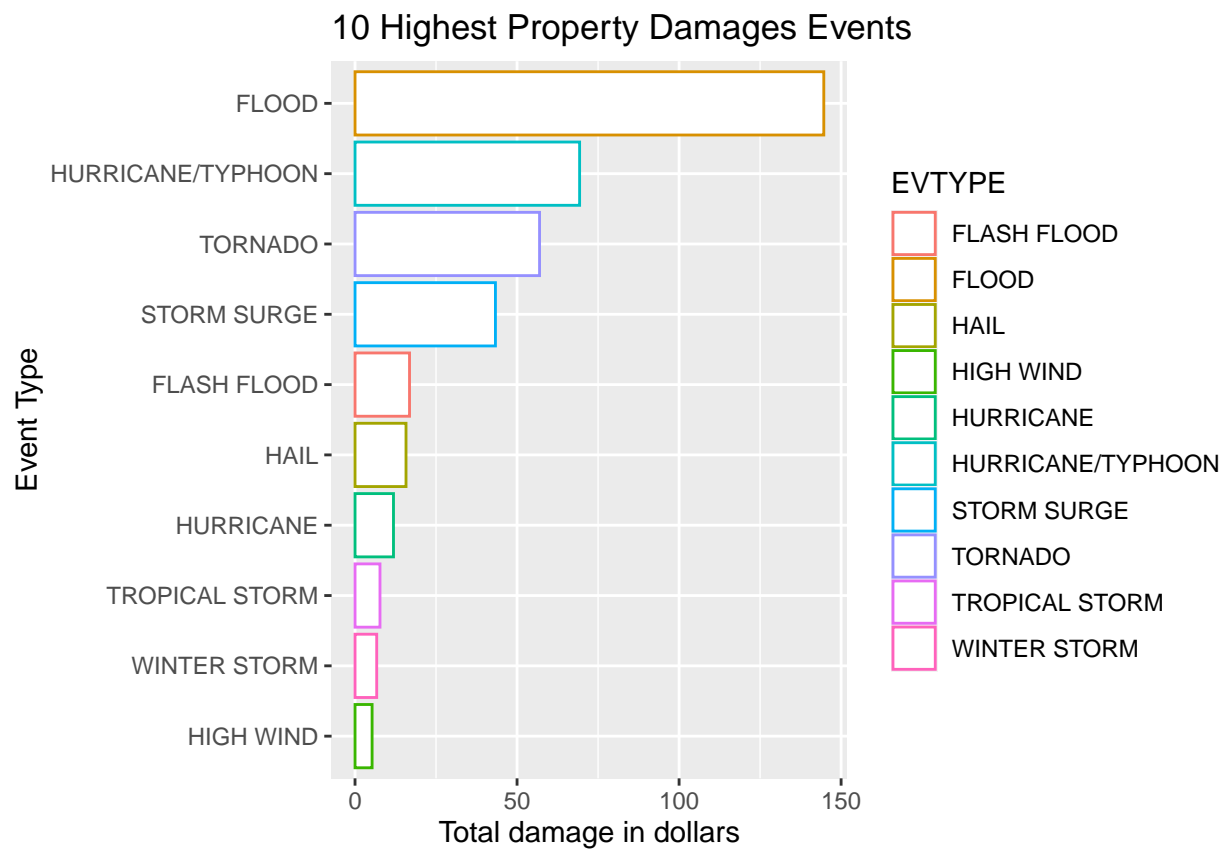
```
aggPdamage <- aggregate(PROPDMGTOTAL ~ EVTYPE, data = stormDTsubset, FUN="sum")
dim(aggPdamage)
```

```
## [1] 985 2
```

```
evPdamage <- aggPdamage[order(-aggPdamage$PROPDMGTOTAL), ][1:10, ]
evPdamage
```

```
##           EVTYPE PROPDMGTOTAL
## 170          FLOOD   144.657710
## 411 HURRICANE/TYPHOON   69.305840
## 834          TORNADO   56.947381
## 670      STORM SURGE   43.323536
## 153      FLASH FLOOD   16.822674
## 244           HAIL   15.735268
## 402          HURRICANE   11.868319
## 848  TROPICAL STORM    7.703891
## 972      WINTER STORM    6.688497
## 359      HIGH WIND    5.270046
```

```
pdamagePlot <- ggplot(data=evPdamage, aes(x=reorder(EVTYPE, PROPDMGTOTAL), y=PROPDMGTOTAL, color=EVTYPE))
pdamagePlot + coord_flip()
```



CROP DAMAGE


```
aggCdamage <- aggregate(CROPDMGTOTAL ~ EVTYPE, data = stormDTsubset, FUN="sum")
dim(aggCdamage)
```

```
## [1] 985 2
```

```
evCdamage <- aggCdamage[order(-aggCdamage$CROPDMGTOTAL), ][1:10, ]
evCdamage
```

```
##           EVTYPE CROPDMGTOTAL
## 95          DROUGHT    13.972566
## 170         FLOOD      5.661968
## 590    RIVER FLOOD      5.029459
## 427    ICE STORM       5.022113
## 244         HAIL       3.025954
## 402    HURRICANE       2.741910
## 411 HURRICANE/TYPHOON  2.607873
## 153    FLASH FLOOD      1.421317
## 140    EXTREME COLD      1.292973
## 212    FROST/FREEZE      1.094086
```

```
cdamagePlot <- ggplot(data=evCdamage, aes(x=reorder(EVTYPE, CROPDMGTOTAL), y=CROPDMGTOTAL, color=EVTYPE))
cdamagePlot + coord_flip()
```

