

Reprocudible_Resarch_Week4

spoltone

29 09 2020

Data load

```
library(knitr)
```

```
## Warning: package 'knitr' was built under R version 4.0.2
```

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

```
## Warning: package 'R.utils' was built under R version 4.0.2
```

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

```
## Warning: package 'R.oo' was built under R version 4.0.2
```

```
## Loading required package: R.methodsS3
```

```
## Warning: package 'R.methodsS3' was built under R version 4.0.2
```

```
## R.methodsS3 v1.8.1 (2020-08-26 16:20:06 UTC) successfully loaded. See  
?R.methodsS3 for help.
```

```
## R.oo v1.24.0 (2020-08-26 16:11:58 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 v2.10.1 (2020-08-26 22:50:31 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, inherits, isOpen, nullfile, parse,
##     warnings

raw_data <- read.csv(file = "repdata_data_StormData.csv", header = TRUE, sep
= ",")
```

```
# summary
```

```
summary(raw_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      END_DATE
## Min.   : 0.000    Length:902297    Length:902297    Length:902297
## 1st Qu.: 0.000    Class :character    Class :character    Class :character
## Median : 0.000    Mode  :character    Mode  :character    Mode  :character
## Mean   : 1.484
## 3rd Qu.: 1.000
## Max.   :3749.000
##
##      END_TIME      COUNTY_END COUNTYENDN      END_RANGE
## Length:902297    Min.   :0    Mode:logical    Min.   : 0.0000
## Class :character    1st Qu.:0    NA's:902297    1st Qu.: 0.0000
## Mode  :character    Median :0
##                      Mean   :0
##                      3rd Qu.:0
##                      Max.   :0
##                      Mean   : 0.9862
##                      3rd Qu.: 0.0000
##                      Max.   :925.0000
##
##      END_AZI      END_LOCATI      LENGTH      WIDTH
## Length:902297    Length:902297    Min.   : 0.0000    Min.   :
```

```

0.000
## Class :character    Class :character    1st Qu.:    0.0000    1st Qu.:
0.000
## Mode  :character    Mode  :character    Median  :    0.0000    Median  :
0.000
##                                     Mean   :    0.2301    Mean   :
7.503
##                                     3rd Qu.:    0.0000    3rd Qu.:
0.000
##                                     Max.    :2315.0000    Max.
:4400.000
##
##          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    Length:902297    Min.      :    0.000    Length:902297
## 1st Qu.:    0.00    Class :character    1st Qu.:    0.000    Class :character
## Median :    0.00    Mode  :character    Median :    0.000    Mode  :character
## Mean   :   12.06                                Mean   :    1.527
## 3rd Qu.:    0.50                                3rd Qu.:    0.000
## Max.   :5000.00                                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
```

```
##
```

```
# check dataset
```

```
head(raw_data)
```

```
## STATE__ BGN_DATE BGN_TIME TIME_ZONE COUNTY COUNTYNAMES STATE  
EVTYPE
```

```
## 1 1 4/18/1950 0:00:00 0130 CST 97 MOBILE AL  
TORNADO
```

```
## 2 1 4/18/1950 0:00:00 0145 CST 3 BALDWIN AL  
TORNADO
```

```
## 3 1 2/20/1951 0:00:00 1600 CST 57 FAYETTE AL  
TORNADO
```

```
## 4 1 6/8/1951 0:00:00 0900 CST 89 MADISON AL  
TORNADO
```

```
## 5 1 11/15/1951 0:00:00 1500 CST 43 CULLMAN AL  
TORNADO
```

```
## 6 1 11/15/1951 0:00:00 2000 CST 77 LAUDERDALE AL  
TORNADO
```

```
## BGN_RANGE BGN_AZI BGN_LOCATI END_DATE END_TIME COUNTY_END COUNTYENDN
```

```
## 1 0 0 0 NA
```

```
## 2 0 0 0 NA
```

```
## 3 0 0 0 NA
```

```
## 4 0 0 0 NA
```

```
## 5 0 0 0 NA
```

```
## 6 0 0 0 NA
```

```
## END_RANGE END_AZI END_LOCATI LENGTH WIDTH F MAG FATALITIES INJURIES
```

```
PROPDMG
```

```
## 1 0 14.0 100 3 0 0 15
```

```
25.0
```

```
## 2 0 2.0 150 2 0 0 0
```

```
2.5
```

```
## 3 0 0.1 123 2 0 0 2
```

```
25.0
```

```
## 4 0 0.0 100 2 0 0 2
```

```
2.5
```

```
## 5 0 0.0 150 2 0 0 2
```

```
2.5
```

```
## 6 0 1.5 177 2 0 0 6
```

```
2.5
```

```
## PROPDMGEXP CROPDGM CROPDGMEXP WFO STATEOFFIC ZONENAMES LATITUDE
```

```
LONGITUDE
```

```
## 1 K 0 3040
```

```
8812
```

```
## 2 K 0 3042
```

```
8755
```

```
## 3 K 0 3340
```

```
8742
```

```
## 4      K      0      3458
8626
## 5      K      0      3412
8642
## 6      K      0      3450
8748
```

```
##  LATITUDE_E LONGITUDE_ REMARKS REFNUM
## 1      3051      8806      1
## 2      0      0      2
## 3      0      0      3
## 4      0      0      4
## 5      0      0      5
## 6      0      0      6
```

```
names(raw_data)
```

```
## [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"
```

RESULTS

QUESTION 1. Across the United States, which types of events (as indicated in the EVTYPE variable) are most harmful with respect to population health?

Reducing the data set to only needed columns

```
main <- c("EVTYPE", "FATALITIES", "INJURIES", "PROPDMG",
"PROPDMGEXP", "CROPDMG", "CROPDMGEXP")
```

```
main_data<-raw_data[main]
```

```
names(main_data)
```

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

Reviewing events that cause the most fatalities and the most injuries

```
Fatalities <- aggregate(FATALITIES ~ EVTYPE, data = main_data, FUN = sum)
Top10_Fatalities <- Fatalities[order(-Fatalities$FATALITIES), ][1:10, ]
Top10_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
```

```
Injuries <- aggregate(INJURIES ~ EVTYPE, data = main_data, FUN = sum)
Top10_Injuries <- Injuries[order(-Injuries$INJURIES), ][1:10, ]
Top10_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
```

Plot of Top 10 Fatalities & Injuries for Weather Event Types

```
par(mfrow=c(1,2),mar=c(10,3,3,2))
barplot(Top10_Fatalities$FATALITIES,names.arg=Top10_Fatalities$EVTYPE,las=2,col="gray",ylab="fatalities",main="Top 10 fatalities")
barplot(Top10_Injuries$INJURIES,names.arg=Top10_Injuries$EVTYPE,las=2,col="gray",ylab="injuries",main="Top 10 Injuries")
```

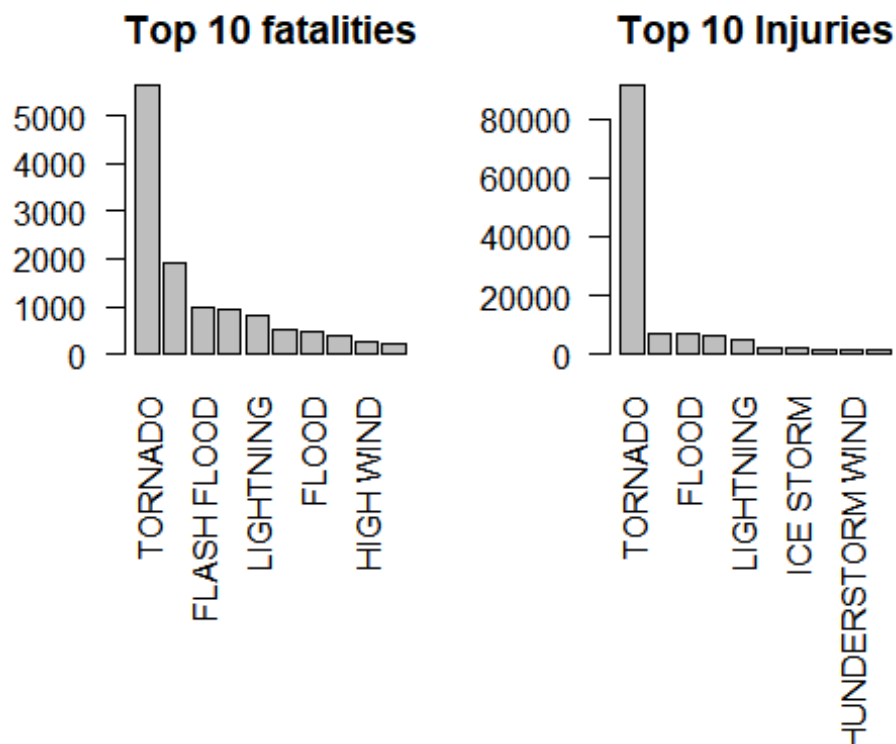


Figure 1: The weather event responsible for the highest fatalities and injuries is the 'Tornado'

QUESTION 2. Across the United States, which types of events have the greatest economic consequences?

Q2.1 Data Exploration & Findings ...

Upon reviewing the column names, the property damage(**PROPDMG**) and crop damage(**CROPDMG**) columns both have another related column titled 'exponents' (i.e - **PROPDMGEXP** and **CROPDMGEXP** respectively).

As a result, let's convert the exponent columns into numeric data for the calculation of total property and crop damages encountered.

```
unique(main_data$PROPDMGEXP)
```

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

```
# Assigning values for the property exponent main_data
main_data$PROPEXP[main_data$PROPDMGEXP == "K"] <- 1000
main_data$PROPEXP[main_data$PROPDMGEXP == "M"] <- 1e+06
main_data$PROPEXP[main_data$PROPDMGEXP == ""] <- 1
main_data$PROPEXP[main_data$PROPDMGEXP == "B"] <- 1e+09
main_data$PROPEXP[main_data$PROPDMGEXP == "m"] <- 1e+06
main_data$PROPEXP[main_data$PROPDMGEXP == "0"] <- 1
main_data$PROPEXP[main_data$PROPDMGEXP == "5"] <- 1e+05
main_data$PROPEXP[main_data$PROPDMGEXP == "6"] <- 1e+06
main_data$PROPEXP[main_data$PROPDMGEXP == "4"] <- 10000
main_data$PROPEXP[main_data$PROPDMGEXP == "2"] <- 100
main_data$PROPEXP[main_data$PROPDMGEXP == "3"] <- 1000
main_data$PROPEXP[main_data$PROPDMGEXP == "h"] <- 100
main_data$PROPEXP[main_data$PROPDMGEXP == "7"] <- 1e+07
main_data$PROPEXP[main_data$PROPDMGEXP == "H"] <- 100
main_data$PROPEXP[main_data$PROPDMGEXP == "1"] <- 10
main_data$PROPEXP[main_data$PROPDMGEXP == "8"] <- 1e+08
```

```
# Assigning '0' to invalid exponent main_data
main_data$PROPEXP[main_data$PROPDMGEXP == "+"] <- 0
main_data$PROPEXP[main_data$PROPDMGEXP == "-"] <- 0
main_data$PROPEXP[main_data$PROPDMGEXP == "?"] <- 0
```

aggregate the property damage by the event type and sort the output it in descending order

```
# property damage summary
propdmg <- aggregate(PROPDMG~EVTYPE,data=main_data,FUN=sum,na.rm=TRUE)
propdmg <- propdmg[with(propdmg,order(-PROPDMG)),]
propdmg <- head(propdmg,10)
print(propdmg)
```

```
##           EVTYPE    PROPDMG
## 834      TORNADO 3212258.2
## 153  FLASH FLOOD 1420124.6
## 856      TSTM WIND 1335965.6
## 170        FLOOD  899938.5
## 760 THUNDERSTORM WIND  876844.2
## 244        HAIL   688693.4
## 464    LIGHTNING  603351.8
## 786 THUNDERSTORM WINDS 446293.2
## 359      HIGH WIND  324731.6
## 972    WINTER STORM 132720.6
```

```
# crop damage summary
crop <- aggregate(CROPDMG~EVTYPE,data=main_data,FUN=sum,na.rm=TRUE)
crop <- crop[with(crop,order(-CROPDMG)),]
crop <- head(crop,10)
print(crop)
```

```
##           EVTYPE    CROPDMG
## 244        HAIL 579596.28
```



```
## 153      FLASH FLOOD 179200.46
## 170      FLOOD 168037.88
## 856      TSTM WIND 109202.60
## 834      TORNADO 100018.52
## 760 THUNDERSTORM WIND 66791.45
## 95      DROUGHT 33898.62
## 786 THUNDERSTORM WINDS 18684.93
## 359      HIGH WIND 17283.21
## 290      HEAVY RAIN 11122.80
```

Plot of Top 10 Property & Crop damages by Weather Event Types

##plot the graph showing the top 10 property and crop damages

```
par(mfrow=c(1,2),mar=c(11,3,3,2))
barplot(propdmg$PROPDMG/(10^9),names.arg=propdmg$EVTYPE,las=2,col="red",ylab="Prop.damage(billions)",main="Top10 Prop.Damages")
barplot(crop$CROPDGMG/(10^9),names.arg=crop$EVTYPE,las=2,col="red",ylab="Crop damage(billions)",main="Top10 Crop.Damages")
```

