# W4 Project2 NOAA storm database

Khalid

7/12/2020

## Your data analysis must address the following questions:

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

 Consider writing your report as if it were to be read by a
 government or municipal manager who might be responsible for
 preparing for severe weather events and will need to
 prioritize resources for different types of events. However,
 there is no need to make any specific recommendations in your
 report.
```

## Load the libraries

```
library(dplyr)
```

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

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

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

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

```
##
## Attaching package: 'data.table'
```

```
## The following objects are masked from 'package:dplyr':
##
##     between, first, last
```

```
library("ggplot2")
```

# Download and extract the data into a dataframe.Then convert to a data.table

```
fileUrl <- "https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2"
download.file(fileUrl, destfile = paste0(getwd(), '/Storm.csv.bz2'), method = 'curl', quiet = T)
storm_DB <- read.csv('Storm.csv.bz2', stringsAsFactors = F)
```

# Converting data.frame to data.table

```
storm_DB <- as.data.table(storm_DB)
```

# Examining DATA

```
names(storm_DB)
```

```
##  [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"
```

# OR

```
colnames(storm_DB)
```

```
##  [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"
```

# Subset the dataset on the parameters that we need and remove unneeded columns.

```
- EVTYPE: event type
- FATALITIES: number of fatalities
- INJURIES: number of injuries
- PROPDMG: property damage (dollars)
- PROPDMGEXP: magnitude of property damage (K = thousands, M =       millions, B = billions)
- CROPDMG: crop damage (dollars)
- CROPDMGEXP: magnitude of crop damage (H = hundreds, K =
thousands, M = millions, B = billions)
```

# Select columns

```
col_select <- c('EVTYPE', 'FATALITIES', 'INJURIES', 'PROPDMG', 'PROPDMGEXP', 'CROPDMG', 'CROPDMG
EXP')

storm1 <- storm_DB[, ..col_select]
summary(storm1)
```

```
##     EVTYPE            FATALITIES          INJURIES            PROPDMG
##  Length:902297      Min.   :  0.0000   Min.   :   0.0000   Min.   :   0.00
##  Class :character   1st Qu.:  0.0000   1st Qu.:   0.0000   1st Qu.:   0.00
##  Mode  :character   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
```

# Find types of events that are most harmful with respect to population health

```
stormDB <- storm1[, 1:3] %>% group_by(EVTYPE) %>% summarise_all(sum)
```

```
summary(stormDB)
```

```
##     EVTYPE            FATALITIES         INJURIES
##  Length:985         Min.   :   0.00   Min.   :    0.0
##  Class :character   1st Qu.:   0.00   1st Qu.:    0.0
##  Mode  :character   Median :   0.00   Median :    0.0
##                     Mean   :  15.38   Mean   :  142.7
##                     3rd Qu.:   0.00   3rd Qu.:    0.0
##                     Max.   :5633.00   Max.   :91346.0
```

# Top 10 events caused death

```
top10Death <- stormDB[order(stormDB$FATALITIES, decreasing = T), ]
head(top10Death)
```

```
## # A tibble: 6 x 3
##   EVTYPE         FATALITIES INJURIES
##   <chr>               <dbl>    <dbl>
## 1 TORNADO              5633    91346
## 2 EXCESSIVE HEAT       1903     6525
## 3 FLASH FLOOD           978     1777
## 4 HEAT                  937     2100
## 5 LIGHTNING             816     5230
## 6 TSTM WIND             504     6957
```

# Top 10 events caused injuries

```
top10inj <- stormDB[order(stormDB$INJURIES, decreasing = T), ]
head(top10inj)
```

```
## # A tibble: 6 x 3
##   EVTYPE         FATALITIES INJURIES
##   <chr>               <dbl>    <dbl>
## 1 TORNADO              5633    91346
## 2 TSTM WIND             504     6957
## 3 FLOOD                 470     6789
## 4 EXCESSIVE HEAT       1903     6525
## 5 LIGHTNING             816     5230
## 6 HEAT                  937     2100
```

# Top 10 events caused both death & injuries

```
stormDB$total <- rowSums(stormDB[, 2:3])
top10both <- stormDB[order(stormDB$total, decreasing = T), ]
top10both[1:10,]
```

```
## # A tibble: 10 x 4
##    EVTYPE            FATALITIES INJURIES total
##    <chr>                  <dbl>    <dbl> <dbl>
##  1 TORNADO                 5633    91346 96979
##  2 EXCESSIVE HEAT          1903     6525  8428
##  3 TSTM WIND                504     6957  7461
##  4 FLOOD                    470     6789  7259
##  5 LIGHTNING                816     5230  6046
##  6 HEAT                     937     2100  3037
##  7 FLASH FLOOD              978     1777  2755
##  8 ICE STORM                 89     1975  2064
##  9 THUNDERSTORM WIND        133     1488  1621
## 10 WINTER STORM             206     1321  1527
```

```
head(top10both)
```

```
## # A tibble: 6 x 4
##   EVTYPE         FATALITIES INJURIES total
##   <chr>               <dbl>    <dbl> <dbl>
## 1 TORNADO              5633    91346 96979
## 2 EXCESSIVE HEAT       1903     6525  8428
## 3 TSTM WIND             504     6957  7461
## 4 FLOOD                 470     6789  7259
## 5 LIGHTNING             816     5230  6046
## 6 HEAT                  937     2100  3037
```

# numbers, characters are all represent factors

```
unique(storm1$PROPDMGEXP)
```

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

# a function to transform

```
getv <- function(EXP_Type) {
  if (EXP_Type %in% c('h', 'H')) {
    return(2)

  } else if (EXP_Type %in% c('k', 'K')) {
    return(3)

  } else if (EXP_Type %in% c('m', 'M')) {
    return(6)

  } else if (EXP_Type %in% c('b', 'B')) {
    return(9)

  } else if (suppressWarnings(!is.na(as.numeric(EXP_Type)))) {

    return(as.numeric(EXP_Type))

  } else {
    return(0)
  }
}
c(10**getv('h'), 10**getv(4), 10**getv('B'), 10**getv('?'))
```

```
## [1] 1e+02 1e+04 1e+09 1e+00
```

# Make a table & Put Result into a Table

```
newST <- storm1[, c(1, 4:7)] %>%
  rowwise() %>%
  mutate(PROP = PROPDMG*10**getv(PROPDMGEXP),
         CROP = CROPDMG*10**getv(CROPDMGEXP))
```

```
head(newST)
```

```
## # A tibble: 6 x 7
## # Rowwise:
##   EVTYPE  PROPDMG PROPDMGEXP CROPDMG CROPDMGEXP  PROP  CROP
##   <chr>     <dbl> <chr>        <dbl> <chr>      <dbl> <dbl>
## 1 TORNADO    25   K                0 ""         25000     0
## 2 TORNADO     2.5 K                0 ""          2500     0
## 3 TORNADO    25   K                0 ""         25000     0
## 4 TORNADO     2.5 K                0 ""          2500     0
## 5 TORNADO     2.5 K                0 ""          2500     0
## 6 TORNADO     2.5 K                0 ""          2500     0
```

```
newST_Sum <- newST[, c(1, 6, 7)] %>%
  group_by(EVTYPE) %>%
  summarise_all(sum)
```

```
summary(newST_Sum)
```

```
##     EVTYPE               PROP               CROP
##  Length:985         Min.   :0.000e+00   Min.   :0.000e+00
##  Class :character   1st Qu.:0.000e+00   1st Qu.:0.000e+00
##  Mode  :character   Median :0.000e+00   Median :0.000e+00
##                     Mean   :4.347e+08   Mean   :4.985e+07
##                     3rd Qu.:5.105e+04   3rd Qu.:0.000e+00
##                     Max.   :1.447e+11   Max.   :1.397e+10
```

# top 10 events that causes the most damages

```
topP <- newST_Sum[order(newST_Sum$PROP, decreasing = T), ]
```

# top 10 events that causes the most crop damages

```
topCrop <- newST_Sum[order(newST_Sum$CROP, decreasing = T), ]
```

# combine the top 10 most damages & crop damages

```
newST_Sum$total <- rowSums(newST_Sum[, 2:3])
combine_top <- newST_Sum[order(newST_Sum$total, decreasing = T), ]
```

```
top10Death[1:10,]
```

```
## # A tibble: 10 x 3
##    EVTYPE         FATALITIES INJURIES
##    <chr>               <dbl>    <dbl>
##  1 TORNADO              5633    91346
##  2 EXCESSIVE HEAT       1903     6525
##  3 FLASH FLOOD           978     1777
##  4 HEAT                  937     2100
##  5 LIGHTNING             816     5230
##  6 TSTM WIND             504     6957
##  7 FLOOD                 470     6789
##  8 RIP CURRENT           368      232
##  9 HIGH WIND             248     1137
## 10 AVALANCHE             224      170
```

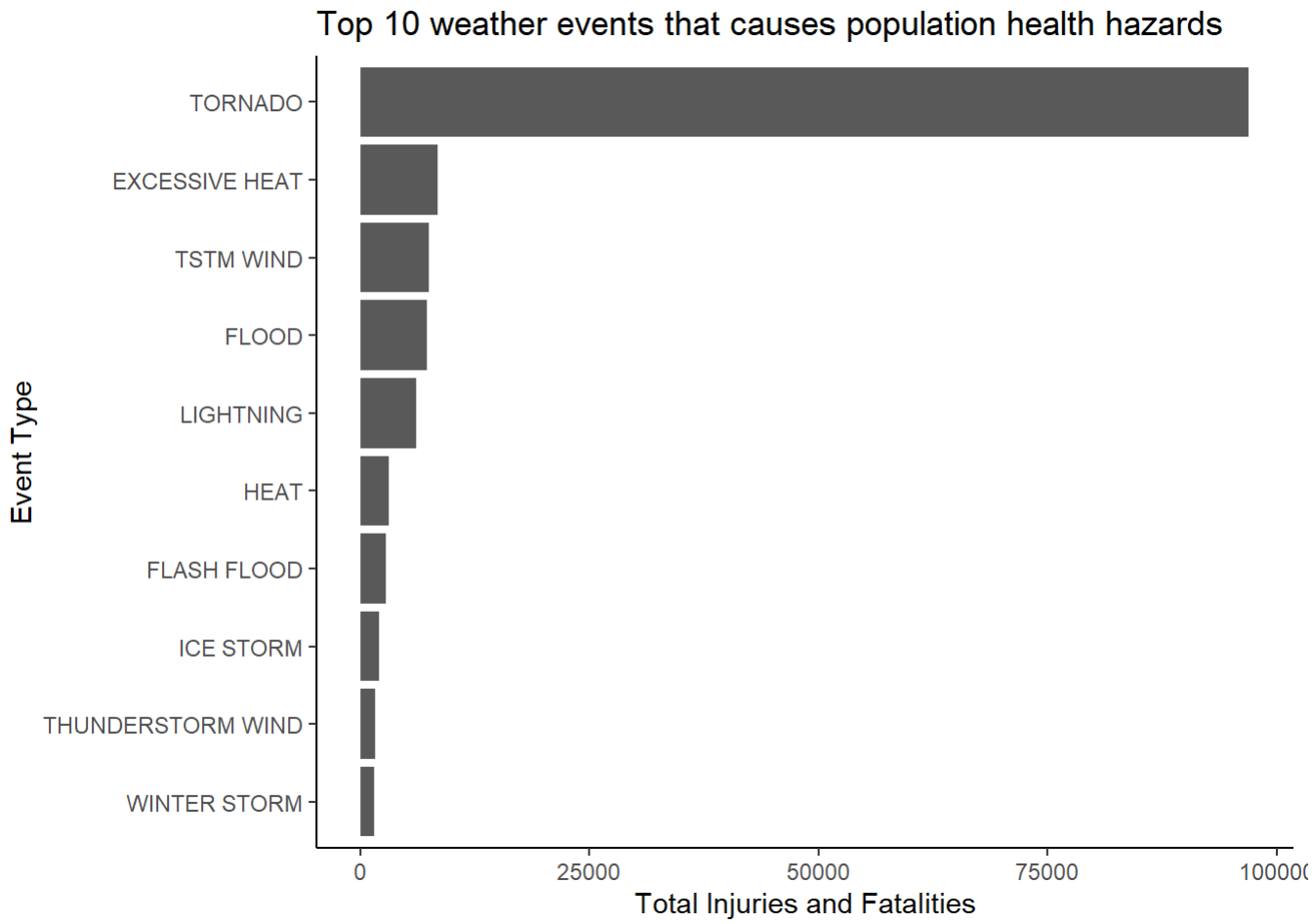
```
top10inj[1:10,]
```

```
## # A tibble: 10 x 3
##    EVTYPE            FATALITIES INJURIES
##    <chr>                  <dbl>    <dbl>
##  1 TORNADO                 5633    91346
##  2 TSTM WIND                504     6957
##  3 FLOOD                    470     6789
##  4 EXCESSIVE HEAT          1903     6525
##  5 LIGHTNING                816     5230
##  6 HEAT                     937     2100
##  7 ICE STORM                 89     1975
##  8 FLASH FLOOD              978     1777
##  9 THUNDERSTORM WIND        133     1488
## 10 HAIL                      15     1361
```

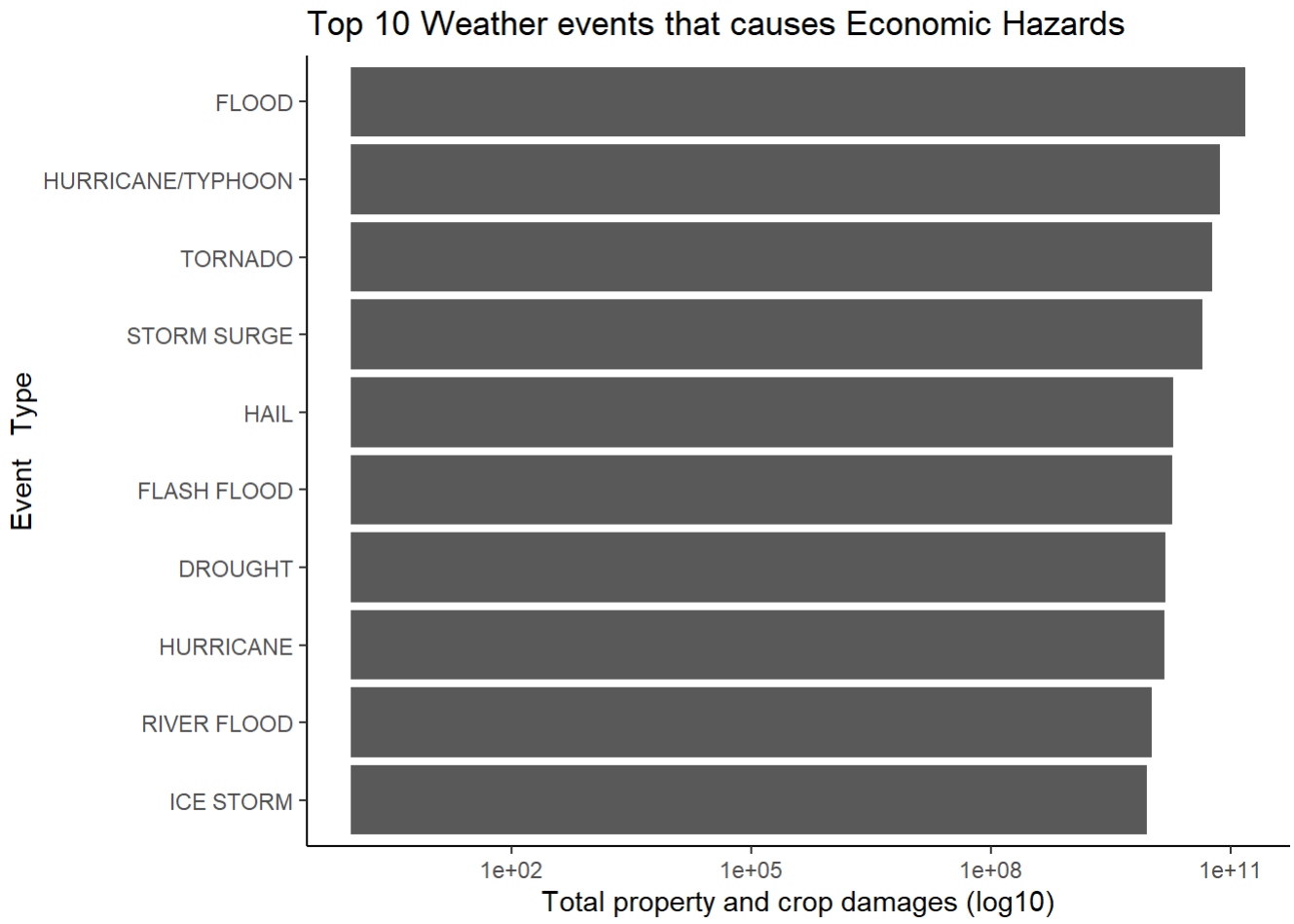
```
top10both[1:10,]
```

```
## # A tibble: 10 x 4
##    EVTYPE            FATALITIES INJURIES total
##    <chr>                  <dbl>    <dbl> <dbl>
##  1 TORNADO                 5633    91346 96979
##  2 EXCESSIVE HEAT          1903     6525  8428
##  3 TSTM WIND                504     6957  7461
##  4 FLOOD                    470     6789  7259
##  5 LIGHTNING                816     5230  6046
##  6 HEAT                     937     2100  3037
##  7 FLASH FLOOD              978     1777  2755
##  8 ICE STORM                 89     1975  2064
##  9 THUNDERSTORM WIND        133     1488  1621
## 10 WINTER STORM             206     1321  1527
```

```
ggplot(data = top10both[1:10,], aes(x = reorder(EVTYPE, total), y = total)) +
  geom_bar(stat = 'Identity') +
  coord_flip() +
  xlab('Event Type') +
  ylab('Total Injuries and Fatalities') +
  ggtitle('Top 10 weather events that causes population health hazards') +
  theme_classic()
```




```
ggplot(data = combine_top[1:10,], aes(x = reorder(EVTYPE, total), y = total)) +
  geom_bar(stat = 'Identity') +
  coord_flip() +
  scale_y_continuous(trans = 'log10') +
  xlab('Event   Type') +
  ylab('Total property and crop damages (log10)') +
  ggtitle('Top 10 Weather events that causes Economic Hazards') +
  theme_classic()
```




```
summary(combine_top)
```

```
##     EVTYPE               PROP                CROP               total          
##  Length:985         Min.   :0.000e+00   Min.   :0.000e+00   Min.   :0.000e+00  
##  Class :character   1st Qu.:0.000e+00   1st Qu.:0.000e+00   1st Qu.:0.000e+00  
##  Mode  :character   Median :0.000e+00   Median :0.000e+00   Median :0.000e+00  
##                     Mean   :4.347e+08   Mean   :4.985e+07   Mean   :4.846e+08  
##                     3rd Qu.:5.105e+04   3rd Qu.:0.000e+00   3rd Qu.:8.500e+04  
##                     Max.   :1.447e+11   Max.   :1.397e+10   Max.   :1.503e+11
```

```
```