Submission

MLG

2022-09-30

Data processing

Data is loaded below and we only select relevant columns:

- *EVTYPE*: Type of disaster (eg. tornado)

- *FATALITIES*: amount of fatalities

- *INJURIES*: amount of injuries

- *PROPDMG*: total property damage

- *PROPDMGEXP*: magnitude of property damage (eg. million)

- *CROPDMG*: total crop damage

- *CROPDMGEXP*: magnitude of crop damage

```
library(dplyr)
library(ggplot2)

df <- read.csv("repdata_data_StormData.csv") %>%
    select(EVTYPE, FATALITIES, INJURIES, PROPDMG, PROPDMGEXP, CROPDMG, CROPDMGEXP)

head(df)
```

```
##     EVTYPE FATALITIES INJURIES PROPDMG PROPDMGEXP CROPDMG CROPDMGEXP
## 1 TORNADO          0       15    25.0          K       0
## 2 TORNADO          0        0     2.5          K       0
## 3 TORNADO          0        2    25.0          K       0
## 4 TORNADO          0        2     2.5          K       0
## 5 TORNADO          0        2     2.5          K       0
## 6 TORNADO          0        6     2.5          K       0
```

Data analysis

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

Harmful events can be defined as the events that caused the most fatalities/injuries. This can be considered over a period (i.e. total sum) or on average (eg. some harmful events do not occur often but cause a lot of damage). We will use both interpretations to answer the question. *df* is transformed below so that we retrieve all the parameters above:

```
harmful_df <- df %>%
    mutate(EVTYPE = trimws(EVTYPE)) %>%
    group_by(EVTYPE) %>%
    summarise(count = n(), sum_injuries = sum(INJURIES), sum_fatalities = sum(FATALITIES),
        mean_injuries = mean(INJURIES), mean_fatalities = mean(FATALITIES)) %>%
    arrange(EVTYPE)
head(harmful_df)
```

```
## # A tibble: 6 x 6
##   EVTYPE              count sum_injuries sum_fatalities mean_injuries mean_fa~1
##   <chr>              <int>        <dbl>          <dbl>         <dbl>     <dbl>
## 1 ?                      1            0              0             0         0
## 2 ABNORMAL WARMTH        4            0              0             0         0
## 3 ABNORMALLY DRY         2            0              0             0         0
## 4 ABNORMALLY WET         1            0              0             0         0
## 5 ACCUMULATED SNOWFALL   4            0              0             0         0
## 6 AGRICULTURAL FREEZE    6            0              0             0         0
## # ... with abbreviated variable name 1: mean_fatalities
```

Most fatalities/injuries

```
harmful_df %>%
    arrange(desc(sum_fatalities)) %>%
    top_n(10)
```

```
## # A tibble: 11 x 6
##    EVTYPE                   count sum_injuries sum_fatalities mean_i~1 mean_~2
##    <chr>                   <int>        <dbl>          <dbl>    <dbl>   <dbl>
## 1  EXTREME HEAT               22          155             96     7.05    4.36
## 2  TORNADOES, TSTM WIND, HAIL  1            0             25     0      25
## 3  RECORD/EXCESSIVE HEAT       3            0             17     0       5.67
## 4  COLD AND SNOW               1            0             14     0      14
## 5  WINTER STORMS               3           17             10     5.67    3.33
## 6  TROPICAL STORM GORDON       1           43              8    43       8
## 7  MARINE MISHAP               2            5              7     2.5     3.5
## 8  HEAT WAVE DROUGHT           1           15              4    15       4
## 9  HIGH WIND/SEAS              1            0              4     0       4
## 10 Heavy surf and wind         1            0              3     0       3
## 11 HIGH WIND AND SEAS          1           20              3    20       3
## # ... with abbreviated variable names 1: mean_injuries, 2: mean_fatalities
```

```
harmful_df %>%
    arrange(desc(sum_injuries)) %>%
    top_n(10)
```

```
## # A tibble: 11 x 6
##    EVTYPE               count sum_injuries sum_fatalities mean_i~1 mean_~2
##    <chr>               <int>        <dbl>          <dbl>    <dbl>   <dbl>
## 1  EXTREME HEAT           22          155             96     7.05    4.36
## 2  TROPICAL STORM GORDON   1           43              8    43       8
## 3  HIGH WIND AND SEAS      1           20              3    20       3
```

```
##  4 WINTER STORMS                 3       17             10     5.67     3.33
##  5 HEAT WAVE DROUGHT             1       15              4    15        4
##  6 MARINE MISHAP                 2        5              7     2.5      3.5
##  7 COLD AND SNOW                 1        0             14     0       14
##  8 Heavy surf and wind           1        0              3     0        3
##  9 HIGH WIND/SEAS                1        0              4     0        4
## 10 RECORD/EXCESSIVE HEAT         3        0             17     0        5.67
## 11 TORNADOES, TSTM WIND, HAIL    1        0             25     0       25
## # ... with abbreviated variable names 1: mean_injuries, 2: mean_fatalities
```

```
harmful_df %>%
    arrange(desc(mean_fatalities)) %>%
    top_n(10)
```

```
## # A tibble: 11 x 6
##    EVTYPE                    count sum_injuries sum_fatalities mean_i~1 mean_~2
##    <chr>                     <int>        <dbl>          <dbl>    <dbl>   <dbl>
##  1 TORNADOES, TSTM WIND, HAIL    1            0             25        0      25
##  2 COLD AND SNOW                 1            0             14        0      14
##  3 TROPICAL STORM GORDON         1           43              8       43       8
##  4 RECORD/EXCESSIVE HEAT         3            0             17        0       5.67
##  5 EXTREME HEAT                 22          155             96        7.05    4.36
##  6 HEAT WAVE DROUGHT             1           15              4       15       4
##  7 HIGH WIND/SEAS                1            0              4        0       4
##  8 MARINE MISHAP                 2            5              7        2.5     3.5
##  9 WINTER STORMS                 3           17             10        5.67    3.33
## 10 Heavy surf and wind           1            0              3        0       3
## 11 HIGH WIND AND SEAS            1           20              3       20       3
## # ... with abbreviated variable names 1: mean_injuries, 2: mean_fatalities
```

```
harmful_df %>%
    arrange(desc(mean_injuries)) %>%
    top_n(10)
```

```
## # A tibble: 11 x 6
##    EVTYPE                    count sum_injuries sum_fatalities mean_i~1 mean_~2
##    <chr>                     <int>        <dbl>          <dbl>    <dbl>   <dbl>
##  1 TROPICAL STORM GORDON         1           43              8       43       8
##  2 HIGH WIND AND SEAS            1           20              3       20       3
##  3 HEAT WAVE DROUGHT             1           15              4       15       4
##  4 EXTREME HEAT                 22          155             96        7.05    4.36
##  5 WINTER STORMS                 3           17             10        5.67    3.33
##  6 MARINE MISHAP                 2            5              7        2.5     3.5
##  7 COLD AND SNOW                 1            0             14        0      14
##  8 Heavy surf and wind           1            0              3        0       3
##  9 HIGH WIND/SEAS                1            0              4        0       4
## 10 RECORD/EXCESSIVE HEAT         3            0             17        0       5.67
## 11 TORNADOES, TSTM WIND, HAIL    1            0             25        0      25
## # ... with abbreviated variable names 1: mean_injuries, 2: mean_fatalities
```