Storm

Ishikawa

2019/7/7

The Analysis the Weather Event Impact to Population Health and Properties

Qestions:

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

From the Storm Events data, answer the questions.

```
library(tidyverse)
library(lubridate)
options(digits=4)
options(warn=-1)
options(width=110)
options(scipen=3)
```

Loading and preprocessing the data

Data from:
National Climatic Data Center Storm Events
(https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2)

First time, Read data file each variavles as character.

```
Storm_wk <- read_csv('repdata_data_StormData.csv.bz2', col_names=T, col_types=str_dup('c',37))

dim(  Storm_wk) %>% cat ('Dimension: ',.,'\n')

## Dimension:  902297 37

names(Storm_wk) %>% cat ('Col Names: ',.,'\n')

## Col Names:  STATE__ BGN_DATE BGN_TIME TIME_ZONE COUNTY COUNTYNAME STATE EVTYPE BGN_RANGE BGN_AZI BGN_LOCA
TI END_DATE END_TIME COUNTY_END COUNTYENDN END_RANGE END_AZI END_LOCATI LENGTH WIDTH F MAG FATALITIES INJURI
ES PROPDMG PROPDMGEXP CROPDMG CROPDMGEXP WFO STATEOFFIC ZONENAMES LATITUDE LONGITUDE LATITUDE_E LONGITUDE_ R
EMARKS REFNUM
```

Relational Valiables for the Questions in the file. EVTYPE - Event Type
FATALITIES
INJURIES
PROPDMG - Property Damage, PROPDMGEXP - Exponents
CROPDMG - Crop Damage, CROPDMGEXP - Exponents

Check the Variable Data characteristics..

```
map(c('EVTYPE', 'FATALITIES', 'INJURIES', 'PROPDMG', 'PROPDMGEXP', 'CROPDMG', 'CROPDMGEXP'),
    function(x){
        cat('\n',x,'---\n')
```

```
        wk1 <- Storm_wk %>% pull(x)
        wk1 %>% is.na() %>% sum() %>% cat('Number of NAs: ',.,'\n')
        wk2 <- wk1 %>% unique()
        wk2 %>% length() %>% cat('Unique Length: ',.,'\n')
        if(length(wk2)<=20){wk2 %>% sort() %>% cat('\n')}
        else               {wk2 %>% sort() %>% head(6) %>% cat('Head: ',.,'\n')
                            wk2 %>% sort() %>% tail(6) %>% cat('Tail: ',.,'\n')}
    }) -> null

##
##  EVTYPE ---
## Number of NAs:  0
## Unique Length:  977
## Head:  ? ABNORMAL WARMTH ABNORMALLY DRY ABNORMALLY WET ACCUMULATED SNOWFALL AGRICULTURAL FREEZE
## Tail:  WINTER WEATHER/MIX WINTERY MIX Wintry mix Wintry Mix WINTRY MIX WND
##
##  FATALITIES ---
## Number of NAs:  0
## Unique Length:  52
## Head:  0.00 1.00 10.00 11.00 114.00 116.00
## Tail:  74.00 75.00 8.00 9.00 90.00 99.00
##
##  INJURIES ---
## Number of NAs:  0
## Unique Length:  200
## Head:  0.00 1.00 10.00 100.00 101.00 102.00
## Tail:  93.00 94.00 95.00 96.00 97.00 98.00
##
##  PROPDMG ---
## Number of NAs:  0
## Unique Length:  1390
## Head:  0.00 0.01 0.02 0.03 0.04 0.05
## Tail:  99.00 99.39 99.97 990.00 995.00 996.00
##
##  PROPDMGEXP ---
## Number of NAs:  465934
## Unique Length:  19
## - ? + 0 1 2 3 4 5 6 7 8 B h H K m M
##
##  CROPDMG ---
## Number of NAs:  0
## Unique Length:  432
## Head:  0.00 0.01 0.02 0.03 0.05 0.10
## Tail:  97.00 975.00 978.00 985.00 99.00 990.00
##
##  CROPDMGEXP ---
## Number of NAs:  618413
## Unique Length:  9
## ? 0 2 B k K m M

Storm_wk <- Storm_wk %>%
         select(BGN_DATE,EVTYPE,FATALITIES,INJURIES,PROPDMG,PROPDMGEXP,CROPDMG,CROPDMGEXP) %>%
         mutate(BGN_DATE   = mdy_hms(BGN_DATE),
                EVTYPE     = str_to_upper(EVTYPE),
                FATALITIES = as.integer(FATALITIES),
                INJURIES   = as.integer(INJURIES),
                PROPDMG    = as.double(PROPDMG),
                PROPDMGEXP = str_to_upper(PROPDMGEXP),
                CROPDMG    = as.double(CROPDMG),
                CROPDMGEXP = str_to_upper(CROPDMGEXP)) %>%
         filter(!str_detect(EVTYPE,'SUMMARY'))
dim(Storm_wk)  # [1] 902221      8

## [1] 902221      8
```

```
Storm_wk %>% head()
```

```
## # A tibble: 6 x 8
##   BGN_DATE            EVTYPE  FATALITIES INJURIES PROPDMG PROPDMGEXP CROPDMG CROPDMGEXP
##   <dttm>              <chr>        <int>    <int>   <dbl> <chr>        <dbl> <chr>
## 1 1950-04-18 00:00:00 TORNADO          0       15    25   K                0 <NA>
## 2 1950-04-18 00:00:00 TORNADO          0        0     2.5 K                0 <NA>
## 3 1951-02-20 00:00:00 TORNADO          0        2    25   K                0 <NA>
## 4 1951-06-08 00:00:00 TORNADO          0        2     2.5 K                0 <NA>
## 5 1951-11-15 00:00:00 TORNADO          0        2     2.5 K                0 <NA>
## 6 1951-11-15 00:00:00 TORNADO          0        6     2.5 K                0 <NA>
```

Check Summary of Data

```
Storm_wk %>% pull(EVTYPE) %>% unique() %>% length()
```

```
## [1] 823
```

```
summary(Storm_wk[,c(3,4,5,7)])
```

```
##    FATALITIES      INJURIES         PROPDMG        CROPDMG
##  Min.   :  0   Min.   :   0.0   Min.   :   0   Min.   :  0.0
##  1st Qu.:  0   1st Qu.:   0.0   1st Qu.:   0   1st Qu.:  0.0
##  Median :  0   Median :   0.0   Median :   0   Median :  0.0
##  Mean   :  0   Mean   :   0.2   Mean   :  12   Mean   :  1.5
##  3rd Qu.:  0   3rd Qu.:   0.0   3rd Qu.:   0   3rd Qu.:  0.0
##  Max.   :583   Max.   :1700.0   Max.   :5000   Max.   :990.0
```

```
Storm_wk %>% pull(PROPDMGEXP) %>% unique() %>% sort()
```

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

```
Storm_wk %>% pull(CROPDMGEXP) %>% unique() %>% sort()
```

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

Clean Event types

There are variations in experssion in adjective for events and event names. So, remove adjectives and be just events. and aggregate event names.

```
Storm <- Storm_wk %>%
        mutate(EVTYPE = str_replace(EVTYPE,'  ',' '),
               EVTYPE = str_replace(EVTYPE,
                        '^BITTER |^RECORD |^EXCESSIVE |^EXTREME |^SMALL |^SEVERE |^HEAVY ',''),
               EVTYPE = str_replace(EVTYPE,
                        '^HIGH |^STRONG |^PROLONG |^UNSEASONABLY |^MODERATE |^DENSE ',''),
               EVTYPE = str_replace(EVTYPE,'^BLIZZARD.*'            ,'BLIZZARD'),
               EVTYPE = str_replace(EVTYPE,'^HAIL.*'                ,'HAIL'),
               EVTYPE = str_replace(EVTYPE,'^HURRICANE.*|^TYPHOON.*','HURRICANE'),
               EVTYPE = str_replace(EVTYPE,'^LIGHTNING.*'           ,'LIGHTNING'),
               EVTYPE = str_replace(EVTYPE,'^THUNDERSTORM.*|^TSTM.*','THUNDERSTORM WIND'),
               EVTYPE = str_replace(EVTYPE,'^COASTAL.*|^BEACH.*'    ,'COASTAL FLOOD'),
               EVTYPE = str_replace(EVTYPE,'^FLASH FLOOD.*'         ,'FLASH FLOOD'),
               EVTYPE = str_replace(EVTYPE,'^FLOOD.*'               ,'FLOOD'),
               EVTYPE = str_replace(EVTYPE,'^URBAN/SML STREAM FLD'  ,'FLOOD'),
               EVTYPE = str_replace(EVTYPE,'^URBAN FLOOD'           ,'FLOOD'),
               EVTYPE = str_replace(EVTYPE,'^RIVER FLOOD.*'         ,'FLOOD'),
               EVTYPE = str_replace(EVTYPE,'^FUNNEL CLOUD.*'        ,'FUNNEL CLOUD'),
               EVTYPE = str_replace(EVTYPE,'^COLD.*'                ,'COLD/WIND CHILL'),
               EVTYPE = str_replace(EVTYPE,'^UN.+ COLD'             ,'COLD/WIND CHILL'),
               EVTYPE = str_replace(EVTYPE,'^HEAT.*|^WARM.*'        ,'HEAT'),
               EVTYPE = str_replace(EVTYPE,'^RAIN.*'                ,'RAIN'),
               EVTYPE = str_replace(EVTYPE,'^MARINE TSTM.*'         ,'MARINE THUNDERSTORM WIND'),
```

```
                EVTYPE = str_replace(EVTYPE,'^MARINE THUNDERSTORM.*' ,'MARINE THUNDERSTORM WIND'),
                EVTYPE = str_replace(EVTYPE,'^SNOW.*'                ,'SNOW'),
                EVTYPE = str_replace(EVTYPE,'^STORM SURGE.*'         ,'STORM SURGE/TIDE'),
                EVTYPE = str_replace(EVTYPE,'^SURF.*'                ,'SURF'),
                EVTYPE = str_replace(EVTYPE,'^TORNADO.*'             ,'TORNADO'),
                EVTYPE = str_replace(EVTYPE,'^WIND.*'                ,'WIND'),
                EVTYPE = str_replace(EVTYPE,'^WATERSPOUT.*'          ,'WATERSPOUT'),
                EVTYPE = str_replace(EVTYPE,'^WILD/FOREST FIRE.*'    ,'WILDFIRE'),
                EVTYPE = str_replace(EVTYPE,'^WINTER WEATHER.*'      ,'WINTER WEATHER')
        )
```

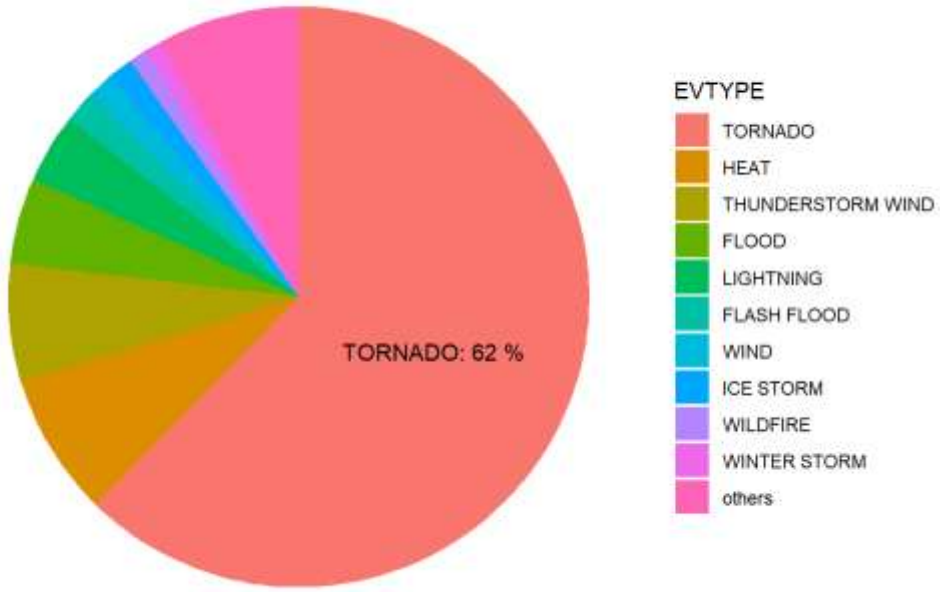
Data Processing for Q1

Which types of events are most harmful with respect to population health?

1) summarize sum of FATALITIES and INJURIES by each event.
2) Sort them in descending order of sum of FATALITIES and INJURIES.
3) Select events by top 10 and make pie chart.

```
event_FAIN <-Storm %>%
        mutate(FATAL_INJU = FATALITIES+INJURIES) %>%
        group_by(EVTYPE) %>%
        summarise(FATAL_INJU=sum(FATAL_INJU)) %>%
        arrange(desc(FATAL_INJU))
event_FAIN <- event_FAIN %>%
        head(10) %>%
        bind_rows(tibble(EVTYPE='others', FATAL_INJU=sum(event_FAIN$FATAL_INJU[11:nrow(event_FAIN)]))) %>%
        mutate(percentage=100*FATAL_INJU/sum(FATAL_INJU))
event_FAIN %>% print(n=Inf)

## # A tibble: 11 x 3
##    EVTYPE            FATAL_INJU percentage
##    <chr>                  <int>      <dbl>
##  1 TORNADO                97022     62.3
##  2 HEAT                   12402      7.97
##  3 THUNDERSTORM WIND      10218      6.56
##  4 FLOOD                   7415      4.76
##  5 LIGHTNING               6049      3.89
##  6 FLASH FLOOD             2803      1.80
##  7 WIND                    2310      1.48
##  8 ICE STORM               2064      1.33
##  9 WILDFIRE                1543      0.991
## 10 WINTER STORM            1527      0.981
## 11 others                 12320      7.91

event_FAIN <- event_FAIN %>%
        mutate(EVTYPE=factor(EVTYPE,levels=event_FAIN$EVTYPE))

ggplot(event_FAIN) +
    geom_bar(aes(x=1,y=FATAL_INJU,fill=EVTYPE),stat='identity') +
    annotate('text', x=1, y=sum(event_FAIN$FATAL_INJU)-event_FAIN$FATAL_INJU[1]/2,
             label=sprintf('%s: %d %%',
                           event_FAIN$EVTYPE[1],
                           as.integer(100*event_FAIN$FATAL_INJU[1]/sum(event_FAIN$FATAL_INJU)))) +
    coord_polar(theta='y') + scale_y_reverse() + theme_void() +
    theme(axis.title.x=element_blank(), axis.title.y=element_blank()) +
    ggtitle('fig1. The top 10 events; Percentage of sum of FATALITIES and INJURIES')
```

fig1. The top 10 events; Percentage of sum of FATALITIES and INJURIES



Data Processing for Q2

Which types of events have the greatest economic consequences?

1) Create ECO_DAMAGE as valiable from PROPDMG, CROPDMG, PROPDMGEXP and CROPDMGEXP. Exponents means Power of 10, and "H" is hundred, "K" is thousand, "M" is Million, "B" is Billion, and others not numeric is 10^0.

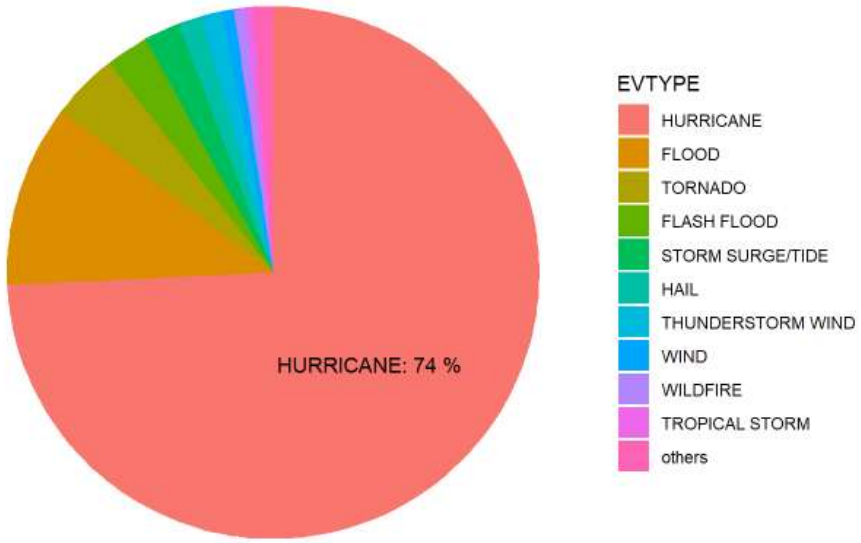
2) summarize ECO_DAMAGE by event.
3) Sort them in descending order of ECO_DAMAGE.
4) Select events bytop 10 ECO_DAMAGE and make pie chart.

```
event_DMG <- Storm %>%
        mutate(PROPDMGEXP=if_else(is.na(PROPDMGEXP),'0',PROPDMGEXP),
               CROPDMGEXP=if_else(is.na(CROPDMGEXP),'0',CROPDMGEXP),
               PROPDMGEXP=recode(PROPDMGEXP, '+'='0','-'='0','?'='0','H'='2','K'='3','M'='6','B'='9'),
               CROPDMGEXP=recode(PROPDMGEXP, '+'='0','-'='0','?'='0','H'='2','K'='3','M'='6','B'='9'),
               ECO_DAMAGE= PROPDMG*(10^as.integer(PROPDMGEXP))+CROPDMG*(10^as.integer(CROPDMGEXP))) %>%
        group_by(EVTYPE) %>%
        summarise(ECO_DAMAGE=sum(ECO_DAMAGE)) %>%
        arrange(desc(ECO_DAMAGE))
event_DMG <- event_DMG %>%
        head(10) %>%
        bind_rows(tibble(EVTYPE='others', ECO_DAMAGE=sum(event_DMG$ECO_DAMAGE[11:nrow(event_DMG)]))) %>%
        mutate(percentage=100*ECO_DAMAGE/sum(ECO_DAMAGE))
event_DMG %>% print(n=Inf)

## # A tibble: 11 x 3
##    EVTYPE            ECO_DAMAGE percentage
##    <chr>                  <dbl>      <dbl>
##  1 HURRICANE           1.64e12     74.3
##  2 FLOOD               2.45e11     11.1
##  3 TORNADO             8.93e10      4.05
##  4 FLASH FLOOD         5.76e10      2.61
```

```
##  5 STORM SURGE/TIDE     4.80e10      2.18
##  6 HAIL                 3.13e10      1.42
##  7 THUNDERSTORM WIND    2.83e10      1.28
##  8 WIND                 1.55e10      0.701
##  9 WILDFIRE             1.50e10      0.682
## 10 TROPICAL STORM       9.65e 9      0.438
## 11 others               2.78e10      1.26
```

```
event_DMG <- event_DMG %>%
        mutate(EVTYPE=factor(EVTYPE,levels=event_DMG$EVTYPE))

ggplot(event_DMG) +
    geom_bar(aes(x=1,y=ECO_DAMAGE,fill=EVTYPE),stat='identity') +
    annotate('text', x=1, y=sum(event_DMG$ECO_DAMAGE)-event_DMG$ECO_DAMAGE[1]/2,
             label=sprintf('%s: %d %%',
                           event_DMG$EVTYPE[1],
                           as.integer(100*event_DMG$ECO_DAMAGE[1]/sum(event_DMG$ECO_DAMAGE)))) +
    coord_polar(theta='y') + scale_y_reverse() + theme_void() +
    theme(axis.title.x=element_blank(), axis.title.y=element_blank()) +
    ggtitle('fig2. The top 10 events; Percentage of sum of Property and Crops Damages')
```

fig2. The top 10 events; Percentage of sum of Property and Crops Damages



The Results

For Question 1

The type of events most harmful with respect to population health is TORNADO. They occupy about 62% of the whole. (Refer fig1)

For Question 2

The types of events the greatest economic consequences is HURRICANE.
They occupy about 74% of the whole. (Refer fig2)