Ranking of national weather events affecting population and economic damage

Gyeong-Geun LEE

2020-12-03

Synopsis

This analysis summarizes the weather events that occurred in the U.S. from 1966 to 2011, resulting in population health and property damage. 10 Major weather events were listed and the percentage were summarized.

Data processing

Basic data processing framework is loaded.

```
library(tidyverse)
library(patchwork)
```

The DB was downloaded and converted into csv file.

```
db_link <- "https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2"
dbfile <- tempfile()
download.file(db_link, dbfile)
raw <- read.csv(dbfile)
unlink(dbfile)
```

All names in the column are converted to lowercase for convenience.

```
df0 <-
  raw %>%
  rename_all(tolower)
head(df0)
```

```
##   state__           bgn_date bgn_time time_zone county countyname state  evtype
## 1       1 4/18/1950 0:00:00     0130       CST     97     MOBILE    AL TORNADO
##   bgn_range bgn_azi bgn_locati end_date end_time county_end countyendn
## 1         0                                               0         NA
##   end_range end_azi end_locati length width f mag fatalities injuries propdmg
## 1         0                        14   100 3   0          0       15      25
##   propdmgexp cropdmg cropdmgexp wfo stateoffic zonenames latitude longitude
## 1          K       0                                         3040      8812
```

```
##   latitude_e longitude_ remarks refnum
## 1       3051       8806              1
##  [ reached 'max' / getOption("max.print") -- omitted 5 rows ]
```

The raw data was glimpsed in order to understand the rough structure of the data.

```
glimpse(df0)
```

```
## Rows: 902,297
## Columns: 37
## $ state__    <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 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 0:0...
## $ bgn_time   <chr> "0130", "0145", "1600", "0900", "1500", "2000", "0100", ...
## $ time_zone  <chr> "CST", "CST", "CST", "CST", "CST", "CST", "CST", "CST", ...
## $ county     <dbl> 97, 3, 57, 89, 43, 77, 9, 123, 125, 57, 43, 9, 73, 49, 1...
## $ countyname <chr> "MOBILE", "BALDWIN", "FAYETTE", "MADISON", "CULLMAN", "L...
## $ state      <chr> "AL", "AL", "AL", "AL", "AL", "AL", "AL", "AL", "AL", "A...
## $ evtype     <chr> "TORNADO", "TORNADO", "TORNADO", "TORNADO", "TORNADO", "...
## $ bgn_range  <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,...
## $ bgn_azi    <chr> "", "", "", "", "", "", "", "", "", "", "", "", "", "", ...
## $ bgn_locati <chr> "", "", "", "", "", "", "", "", "", "", "", "", "", "", ...
## $ end_date   <chr> "", "", "", "", "", "", "", "", "", "", "", "", "", "", ...
## $ end_time   <chr> "", "", "", "", "", "", "", "", "", "", "", "", "", "", ...
## $ county_end <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,...
## $ countyendn <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
## $ end_range  <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,...
## $ end_azi    <chr> "", "", "", "", "", "", "", "", "", "", "", "", "", "", ...
## $ end_locati <chr> "", "", "", "", "", "", "", "", "", "", "", "", "", "", ...
## $ length     <dbl> 14.0, 2.0, 0.1, 0.0, 0.0, 1.5, 1.5, 0.0, 3.3, 2.3, 1.3, ...
## $ width      <dbl> 100, 150, 123, 100, 150, 177, 33, 33, 100, 100, 400, 400...
## $ f          <int> 3, 2, 2, 2, 2, 2, 2, 1, 3, 3, 1, 1, 3, 3, 3, 4, 1, 1, 1,...
## $ mag        <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,...
## $ fatalities <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 4, 0, 0, 0,...
## $ injuries   <dbl> 15, 0, 2, 2, 2, 6, 1, 0, 14, 0, 3, 3, 26, 12, 6, 50, 2, ...
## $ propdmg    <dbl> 25.0, 2.5, 25.0, 2.5, 2.5, 2.5, 2.5, 2.5, 25.0, 25.0, 2....
## $ propdmgexp <chr> "K", "K", "K", "K", "K", "K", "K", "K", "K", "K", "M", "...
## $ cropdmg    <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,...
## $ cropdmgexp <chr> "", "", "", "", "", "", "", "", "", "", "", "", "", "", ...
## $ wfo        <chr> "", "", "", "", "", "", "", "", "", "", "", "", "", "", ...
## $ stateoffic <chr> "", "", "", "", "", "", "", "", "", "", "", "", "", "", ...
## $ zonenames  <chr> "", "", "", "", "", "", "", "", "", "", "", "", "", "", ...
## $ latitude   <dbl> 3040, 3042, 3340, 3458, 3412, 3450, 3405, 3255, 3334, 33...
## $ longitude  <dbl> 8812, 8755, 8742, 8626, 8642, 8748, 8631, 8558, 8740, 87...
## $ latitude_e <dbl> 3051, 0, 0, 0, 0, 0, 0, 0, 3336, 3337, 3402, 3404, 0, 34...
## $ longitude_ <dbl> 8806, 0, 0, 0, 0, 0, 0, 0, 8738, 8737, 8644, 8640, 0, 85...
## $ remarks    <chr> "", "", "", "", "", "", "", "", "", "", "", "", "", "", ...
## $ refnum     <dbl> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 1...
```

Important features were seleted to analyze.

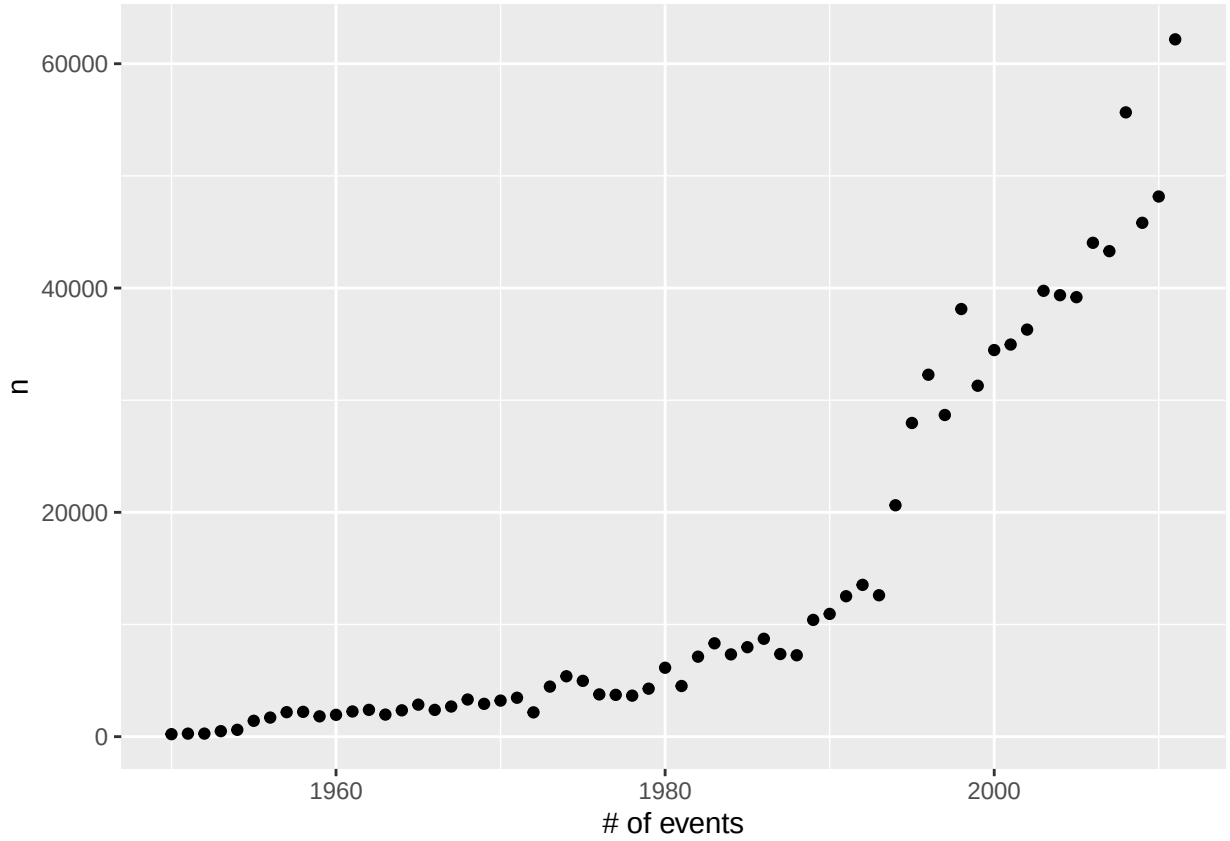
- `state`: state
- `bgn_date`" begin date
- `evtype`: event type

- `fatalities`
- `injuries`
- `propdmg`: property damage
- `propdmgexp`: exponent of property damage
- `cropdmg`: crop damage
- `cropdmgexp`: exponent of crop damage

```
# selected features
df1 <-
  df0 %>%
  select(
    state,
    bgn_date,
    evtype,
    fatalities, injuries,
    propdmg, propdmgexp, cropdmg, cropdmgexp
  ) %>%
  mutate(event_date = lubridate::mdy_hms(bgn_date)) %>%
  relocate(evtype, state, event_date) %>%
  select(-bgn_date)
```

In order to cut the limit of the year, the event number was shown according to year.

```
df1 %>%
  mutate(year = lubridate::year(event_date)) %>%
  count(year) %>%
  ggplot(aes(year, n)) +
  geom_point() +
  labs(x = "# of events")
```



Since the previous data is less reliable, only data after 1990 were considered.

```
df1 <-
  df1 %>%
  filter(lubridate::year(event_date) >= 1990)
```

The NA data were removed.

```
sum(!complete.cases(df1))
```

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

The exponent of property damage and crop damage were anlayzed

```
table(df1$propdmgexp)
```

```
##
##             -      ?      +      0      1      2      3      4      5      6
## 346265      1      8      5    216     25     13      4      4     28      4
##      7      8      B      h      H      K      m      M
##      5      1     40      1      6 396151      7   8956
```

```
table(df1$cropdmgexp)
```

```
##
##             ?      0      2      B      k      K      m      M
## 467856      7     19      1      9     21 281832      1   1994
```

The exponent was converted into numbers, and incoporated to mantissa.

```
convert_exp <- function(s) {
  #s <- as.character()
  r <-
    case_when(
      s %in% c("", "-", "?", "+", "0", "h", "H") ~ "0",
      s %in% c("k", "K") ~ "3",
      s %in% c("m", "M") ~ "6",
      s %in% c("b", "B") ~ "9",
      s %in% c("1", "2", "3", "4", "5", "6", "7", "8") ~ s,
      TRUE ~ "0"
    )
  10^as.numeric(r)
}

df1 <-
  df1 %>%
  mutate(
    propdmgexp2 = convert_exp(propdmgexp),
    cropdmgexp2 = convert_exp(cropdmgexp),
    prop = propdmg * propdmgexp2,           # finally converted value
    crop = cropdmg * cropdmgexp2            # finally converted value
  ) %>%
  relocate(evtype:injuries, prop, crop)
df1
```

```
##      evtype state event_date fatalities injuries prop crop propdmg propdmgexp
## 1      HAIL    AL 1990-01-05          0        0    0    0       0
## 2 TSTM WIND    AL 1990-01-20          0        0    0    0       0
## 3 TSTM WIND    AL 1990-01-20          0        0    0    0       0
##   cropdmg cropdmgexp propdmgexp2 cropdmgexp2
## 1       0                     1           1
## 2       0                     1           1
## 3       0                     1           1
##  [ reached 'max' / getOption("max.print") -- omitted 751737 rows ]
```

Results

events vs population health

The total was summarized for each event type.

```
df1_population <-
  df1 %>%
  group_by(evtype) %>%
  summarize(
    fatalities = sum(fatalities),
    injuries = sum(injuries),
    .groups = "drop_last"
  )
df1_population
```

```
## # A tibble: 985 x 3
##    evtype                  fatalities injuries
##    <chr>                        <dbl>    <dbl>
##  1 "   HIGH SURF ADVISORY"          0        0
##  2 " COASTAL FLOOD"                 0        0
##  3 " FLASH FLOOD"                   0        0
##  4 " LIGHTNING"                     0        0
##  5 " TSTM WIND"                     0        0
##  6 " TSTM WIND (G45)"               0        0
##  7 " WATERSPOUT"                    0        0
##  8 " WIND"                          0        0
##  9 "?"                              0        0
## 10 "ABNORMAL WARMTH"                0        0
## # ... with 975 more rows
```

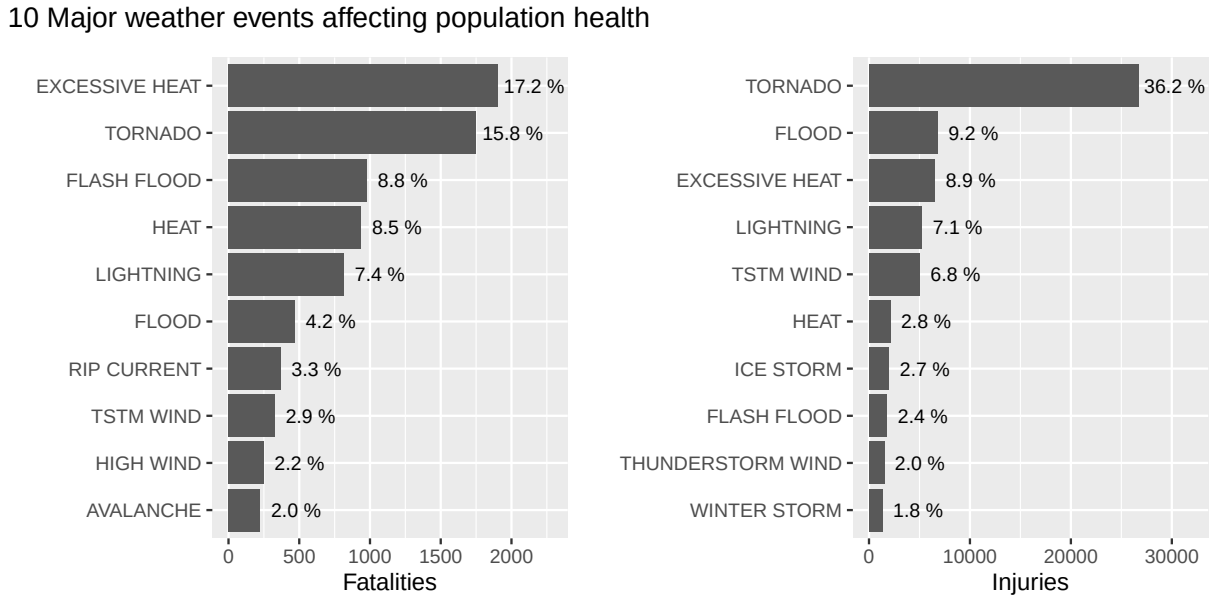
The fatalities and injuries were expressed in a figure.

```
g1 <-
  df1_population %>%
  mutate(prop = fatalities / sum(fatalities)) %>%
  arrange(desc(fatalities)) %>%
  slice(1:10) %>%
  ggplot(aes(reorder(evtype, fatalities), fatalities)) +
  geom_bar(stat = "identity") +
  geom_text(
    aes(label = paste(format(round(prop * 100, 1), nsmall = 1), "%")),
    hjust = -0.1,
    size = 3
  ) +
  scale_y_continuous(limits = c(0, max(df1_population$fatalities)*1.2)) +
  coord_flip() +
  labs(x = "", y = "Fatalities")

g2 <-
  df1_population %>%
  mutate(prop = injuries / sum(injuries)) %>%
  arrange(desc(injuries)) %>%
  slice(1:10) %>%
  ggplot(aes(reorder(evtype, injuries), injuries)) +
  geom_bar(stat = "identity") +
  geom_text(
    aes(label = paste(format(round(prop * 100, 1), nsmall = 1), "%")),
    hjust = -0.1,
```

```
    size = 3
  ) +
  scale_y_continuous(limits = c(0, max(df1_population$injuries)*1.2)) +
  coord_flip() +
  labs(x = "", y = "Injuries")

g1 + g2 +
  plot_annotation(title = "10 Major weather events affecting population health")
```



events vs economic consequences

```
df1_economics <-
  df1 %>%
  group_by(evtype) %>%
  summarize(
    prop_sum = sum(prop),
    crop_sum = sum(crop),
    .groups = "drop_last"
  )
df1_economics
```

```
## # A tibble: 985 x 3
##    evtype                  prop_sum crop_sum
##    <chr>                      <dbl>    <dbl>
##  1 "   HIGH SURF ADVISORY"   200000        0
##  2 " COASTAL FLOOD"              0        0
##  3 " FLASH FLOOD"            50000        0
##  4 " LIGHTNING"                  0        0
##  5 " TSTM WIND"            8100000        0
##  6 " TSTM WIND (G45)"         8000        0
```

```
##  7 " WATERSPOUT"                  0        0
##  8 " WIND"                        0        0
##  9 "?"                         5000        0
## 10 "ABNORMAL WARMTH"              0        0
## # ... with 975 more rows
```

The population damage and crop damage were expressed in a figure.

```
temp <-
  df1_economics %>%
  mutate(
    dmg = prop_sum / 1e6,
    prop = dmg / sum(dmg)
  )

g1 <-
  temp %>%
  arrange(desc(prop_sum)) %>%
  slice(1:10) %>%
  ggplot(aes(reorder(evtype, dmg), dmg)) +
  geom_bar(stat = "identity") +
  geom_text(
    aes(label = paste(format(round(prop * 100, 1), nsmall = 1), "%")),
    hjust = -0.1,
    size = 3
  ) +
  scale_y_continuous(
    limits = c(0, max(temp$dmg)*1.2),
  ) +
  coord_flip() +
  labs(x = "", y = "Property damage $ (×1000)")

temp <-
  df1_economics %>%
  mutate(
    dmg = crop_sum / 1e6,
    prop = dmg / sum(dmg)
  )

g2 <-
  temp %>%
  arrange(desc(crop_sum)) %>%
  slice(1:10) %>%
  ggplot(aes(reorder(evtype, dmg), dmg)) +
  geom_bar(stat = "identity") +
  geom_text(
    aes(label = paste(format(round(prop * 100, 1), nsmall = 1), "%")),
    hjust = -0.1,
    size = 3
  ) +
  scale_y_continuous(
    limits = c(0, max(temp$dmg)*1.2),
  ) +
  coord_flip() +
```

```
  labs(x = "", y = "Crop damage $ (×1000)")

g1 + g2 +
  plot_annotation(title = "10 Major weather events affecting economic consequences")
```