Analysis on severe weather events in the USA

Reproducible Research: Peer-graded Assignment 2

Luis Escobar Sawa

27/10/2022

Analysis on severe weather events in the USA

The goal of this study will be to try to address the following questions:

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

Data Processing

We are going to use the tidyverse library which is a collection of useful libraries like, dplyr, tibble, ggplot2 and readr, among others.

```
library(tidyverse)
```

First we download the database if not already present in the repository.

```
dataDir <- "database/"
dataFile <- paste0(dataDir, "NOAA_stormdatabase.csv.bz2")
dataURL <- "https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2"

if(!dir.exists(dataDir)){
    dir.create(dataDir)
}

if(!file.exists(dataFile)){
    download.file(dataURL, dataFile)
}
```

We read the first observation to get the information on the columns present in the dataset.

```
data <- read_csv(dataFile, n_max = 1, show_col_types = FALSE)
spec(data)
```

```
## cols(
##   STATE__ = col_double(),
##   BGN_DATE = col_character(),
```

```
##   BGN_TIME = col_character(),
##   TIME_ZONE = col_character(),
##   COUNTY = col_double(),
##   COUNTYNAME = col_character(),
##   STATE = col_character(),
##   EVTYPE = col_character(),
##   BGN_RANGE = col_double(),
##   BGN_AZI = col_logical(),
##   BGN_LOCATI = col_logical(),
##   END_DATE = col_logical(),
##   END_TIME = col_logical(),
##   COUNTY_END = col_double(),
##   COUNTYENDN = col_logical(),
##   END_RANGE = col_double(),
##   END_AZI = col_logical(),
##   END_LOCATI = col_logical(),
##   LENGTH = col_double(),
##   WIDTH = col_double(),
##   F = col_double(),
##   MAG = col_double(),
##   FATALITIES = col_double(),
##   INJURIES = col_double(),
##   PROPDMG = col_double(),
##   PROPDMGEXP = col_character(),
##   CROPDMG = col_double(),
##   CROPDMGEXP = col_logical(),
##   WFO = col_logical(),
##   STATEOFFIC = col_logical(),
##   ZONENAMES = col_logical(),
##   LATITUDE = col_double(),
##   LONGITUDE = col_double(),
##   LATITUDE_E = col_double(),
##   LONGITUDE_ = col_double(),
##   REMARKS = col_logical(),
##   REFNUM = col_double()
## )
```

I have decided to only keep the following columns:

- BGN_DATE: The date of the start of the observation. While not completely relevant for this study, it will prove useful in some decisions.
- EVTYPE: The type of the event observed. It is the main variable that we will be looking at.
- FATALITIES: The number of fatalities caused by each event and is, of course, a key variable to investigate impact on population health.
- INJURIES: The number of injuries caused by each event. This is the second variable we will use with regard to population health.
- PROPDMG: The amount of property damage caused by the event in USD.
- PROPDMGEXP: The magnitude of the property damage in the previous column.
- CROPDMG: The amount of damage to crops caused by the event in USD.
- CROPDMGEXP: The magnitude of the damage to crops in the previous column.

```
data <- read_csv(dataFile, col_types = cols_only(
  BGN_DATE = col_date("%m/%d/%Y 0:00:00"),
```

```r
  EVTYPE = col_character(),
  FATALITIES = col_double(),
  INJURIES = col_double(),
  PROPDMG = col_double(),
  PROPDMGEXP = col_character(),
  CROPDMG = col_double(),
  CROPDMGEXP = col_character()))
```

```r
subdata <- subset(data, PROPDMGEXP %in% c("+","0","5","6","?","4","2","3","h","7","H","-","1","8"))
```

```r
data$EVTYPE <- sapply(data$EVTYPE, str_to_lower)
```

Explore the data

```r
summary <- data %>% group_by(EVTYPE) %>% summarise(FATALITIES=sum(FATALITIES), INJURIES=sum(INJURIES), 
```

```r
fatalities <- summary$FATALITIES
injuries <- summary$INJURIES
propdmg <- summary$PROPDMG
cropdmg <- summary$CROPDMG
```

```r
quantile(fatalities, probs = seq(0.85, 1, .01))
```

```
##      85%      86%      87%      88%      89%      90%      91%      92%      93%      94%
##     1.00     1.00     1.00     2.00     2.21     3.00     4.00     5.00     7.00    11.00
##      95%      96%      97%      98%      99%     100%
##    17.00    30.76    62.66   103.22   226.64  5633.00
```

```r
unique(data[data$FATALITIES>0,]$EVTYPE)
```

```
##    [1] "tornado"                    "tstm wind"
##    [3] "hail"                       "winter storm"
##    [5] "hurricane opal/high winds"  "dense fog"
##    [7] "rip current"                "heat"
##    [9] "lightning"                  "cold"
##   [11] "flooding"                   "heavy rain"
##   [13] "flash flood"                "extreme cold"
##   [15] "thunderstorm winds"         "marine mishap"
##   [17] "high wind/seas"             "high seas"
##   [19] "dust storm"                 "sleet"
##   [21] "flood"                      "excessive heat"
##   [23] "gusty winds"                "high winds"
##   [25] "high surf"                  "wild fires"
##   [27] "winter storm high winds"    "winter storms"
##   [29] "thunderstorm wind"          "flood/flash flood"
##   [31] "heavy snow"                 "ice storm"
##   [33] "heat wave"                  "unseasonably warm"
##   [35] "strong winds"               "hurricane erin"
##   [37] "wind"                       "hurricane opal"
```

```
##  [39] "tornadoes, tstm wind, hail"    "tropical storm gordon"
##  [41] "waterspout"                      "storm surge"
##  [43] "blizzard"                        "freezing rain/snow"
##  [45] "thundersnow"                     "high wind"
##  [47] "flash flooding"                  "freezing rain"
##  [49] "fog"                             "snow and ice"
##  [51] "wind storm"                      "ice"
##  [53] "urban and small stream floodin" "waterspout/tornado"
##  [55] "extreme heat"                    "freeze"
##  [57] "lightning."                      "thundertorm winds"
##  [59] "cold wave"                       "flood/river flood"
##  [61] "river flood"                     "avalance"
##  [63] "heavy snow and high winds"       "rip currents/heavy surf"
##  [65] "fog and cold temperatures"       "heavy surf"
##  [67] "icy roads"                       "snow"
##  [69] "rip currents"                    "hurricane felix"
##  [71] "blowing snow"                    "rain/wind"
##  [73] "heat wave drought"               "heat waves"
##  [75] "unseasonably warm and dry"       "unseasonably cold"
##  [77] "record/excessive heat"           "thunderstorm wind g52"
##  [79] "high waves"                      "low temperature"
##  [81] "hypothermia"                     "cold/winds"
##  [83] "record cold"                     "snow/ bitter cold"
##  [85] "cold weather"                    "rapidly rising water"
##  [87] "high winds/snow"                 "flash flood/flood"
##  [89] "excessive rainfall"              "flash flooding/flood"
##  [91] "glaze"                           "landslide"
##  [93] "high wind and seas"              "river flooding"
##  [95] "minor flooding"                  "drought/excessive heat"
##  [97] "heavy seas"                      "avalanche"
##  [99] "flood & heavy rain"              "flash floods"
## [101] "tropical storm"                  "urban/sml stream fld"
## [103] "rough surf"                      "marine accident"
## [105] "dry microburst"                  "winds"
## [107] "coastal storm"                   "hurricane"
## [109] "extended cold"                   "extreme windchill"
## [111] "whirlwind"                       "mixed precip"
## [113] "freezing spray"                  "mudslides"
## [115] "strong wind"                     "cold temperature"
## [117] "coastal flooding"                "cold and snow"
## [119] "rain/snow"                       "hypothermia/exposure"
## [121] "black ice"                       "coastalstorm"
## [123] "freezing drizzle"                "frost"
## [125] "wild/forest fire"                "snow squalls"
## [127] "mudslide"                        "heavy surf and wind"
## [129] "landslides"                      "high swells"
## [131] "tstm wind/hail"                  "tstm wind (g35)"
## [133] "snow squall"                     "hyperthermia/exposure"
## [135] "record heat"                     "gusty wind"
## [137] "wintry mix"                      "rough seas"
## [139] "thunderstorm wind (g40)"         "high water"
## [141] "light snow"                      "thunderstorm"
## [143] "falling snow/ice"                "ice on road"
## [145] "drowning"                        "extreme cold/wind chill"
```

```
## [147]  "hurricane/typhoon"              "wildfire"
## [149]  "heavy surf/high surf"           "winter weather/mix"
## [151]  "dust devil"                     "marine tstm wind"
## [153]  "winter weather"                 "cold/wind chill"
## [155]  "marine thunderstorm wind"       "marine strong wind"
## [157]  "coastal flood"                  "storm surge/tide"
## [159]  "marine high wind"               "tsunami"
```

Astronomical Low Tide Z Avalanche Z Blizzard Z Coastal Flood Z Cold/Wind Chill Z Debris Flow C Dense Fog Z Dense Smoke Z Drought Z Dust Devil C Dust Storm Z Excessive Heat Z Extreme Cold/Wind Chill Z Flash Flood C Flood C Frost/Freeze Z Funnel Cloud C Freezing Fog Z Hail C Heat Z Heavy Rain C Heavy Snow Z High Surf Z High Wind Z Hurricane (Typhoon) Z Ice Storm Z Lake-Effect Snow Z Lakeshore Flood Z Lightning C Marine Hail M Marine High Wind M Marine Strong Wind M Marine Thunderstorm Wind M Rip Current Z Seiche Z Sleet Z Storm Surge/Tide Z Strong Wind Z Thunderstorm Wind C Tornado C Tropical Depression Z Tropical Storm Z Tsunami Z Volcanic Ash Z Waterspout M Wildfire Z Winter Storm Z Winter Weather Z

Results

You can also embed plots, for example:

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.