Exploring NOAA Storm Database to Determine Most Impactful Severe Weather Events

Troy Huffman

2/13/2022

Synopsis

This data analysis explores the U.S. National Oceanic and Atmospheric Administration's (NOAA) storm database between 1950 and 2011 to determine which severe weather event is most harmful to the human population and its economy.Analysis detemines tornados and excessive heat are most harmful to the human population and floods and hurricanes are most harmful to its economy.

Data Processing

R code with comments will be displayed to inform the reader of the processes used to perform each step of analysis.

```
knitr::opts_chunk$set(echo = TRUE)
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(tidyr)
library(ggplot2)
```

Loading the NOAA storm data file and storing it in data frame format for analysis. Data file uploaded to Coursera Lab Sandbox.

```
storm_data <- read.csv("repdata_data_StormData.csv.bz2")
```

Selecting the following variables in the data that are relevant to our questions:

EVTYPE: Event Types FATALITIES: Fatalities Caused INJURIES: Injuries Caused PROPDMG: Property Damage Caused PROPDMGEXP: Property Damage Magnitude CROPDMG: Crop Damage Caused CROPDMGEXP: Crop Damage Magnitude

```
storm_data_variables <- c("EVTYPE","FATALITIES","INJURIES","PROPDMG","PROPDMGEXP" ,"CROPDMG","CROPDMGEX
new_storm_data <-select(storm_data, all_of(storm_data_variables))
```

Some EVTYPES represent the same weather event, but are recorded as different terms. The following code ensures all weather events are represented with one matching EVTYPE.

```
new_storm_data$EVTYPE[grepl("FLOOD", new_storm_data$EVTYPE, ignore.case = TRUE)] <- "FLOOD"
new_storm_data$EVTYPE[grepl("TORNADO", new_storm_data$EVTYPE, ignore.case = TRUE)] <- "TORNADO"
new_storm_data$EVTYPE[grepl("TSTM|THUNDERSTORM", new_storm_data$EVTYPE, ignore.case = TRUE)] <- "TSTM"
new_storm_data$EVTYPE[grepl("TROPICAL|STORM", new_storm_data$EVTYPE, ignore.case = TRUE)] <- "STORM"
new_storm_data$EVTYPE[grepl("HURRICANE", new_storm_data$EVTYPE, ignore.case = TRUE)] <- "HURRICANE"
new_storm_data$EVTYPE[grepl("ICE|SNOW|FROST|SLEET", new_storm_data$EVTYPE, ignore.case = TRUE)] <- "SNO
new_storm_data$EVTYPE[grepl("FOG", new_storm_data$EVTYPE, ignore.case = TRUE)] <- "FOG"
new_storm_data$EVTYPE[grepl("COLD|WINDCHILL|FREEZE|WINTER", new_storm_data$EVTYPE, ignore.case = TRUE)]
new_storm_data$EVTYPE[grepl("HEAT|WARM|HOT", new_storm_data$EVTYPE, ignore.case = TRUE)] <- "HEAT"
new_storm_data$EVTYPE[grepl("CLOUD|FUNNEL", new_storm_data$EVTYPE, ignore.case = TRUE)] <- "CLOUD"
new_storm_data$EVTYPE[grepl("HAIL", new_storm_data$EVTYPE, ignore.case = TRUE)] <- "HAIL"
new_storm_data$EVTYPE[grepl("DROUGHT|DRY", new_storm_data$EVTYPE, ignore.case = TRUE)] <- "DROUGHT"
new_storm_data$EVTYPE[grepl("LIGHTNING", new_storm_data$EVTYPE, ignore.case = TRUE)] <- "LIGHTNING"
new_storm_data$EVTYPE[grepl("FIRE", new_storm_data$EVTYPE, ignore.case = TRUE)] <- "FIRE"
new_storm_data$EVTYPE[grepl("RAIN|SHOWER", new_storm_data$EVTYPE, ignore.case = TRUE)] <- "RAIN"
new_storm_data$EVTYPE[grepl("WATERSPOUT", new_storm_data$EVTYPE, ignore.case = TRUE)] <- "WATERSPOUT"
new_storm_data$EVTYPE[grepl("SURF", new_storm_data$EVTYPE, ignore.case = TRUE)] <- "SURF"
new_storm_data$EVTYPE[grepl("CURRENT", new_storm_data$EVTYPE, ignore.case = TRUE)] <- "CURRENT"
new_storm_data$EVTYPE[grepl("WIND|MICROBURST", new_storm_data$EVTYPE, ignore.case = TRUE)] <- "WIND"
new_storm_data$EVTYPE[grepl("BLIZZARD", new_storm_data$EVTYPE, ignore.case = TRUE)] <- "BLIZZARD"
new_storm_data$EVTYPE[grepl("SLIDE", new_storm_data$EVTYPE, ignore.case = TRUE)] <- "LANDSLIDE"
new_storm_data$EVTYPE[grepl("DUST", new_storm_data$EVTYPE, ignore.case = TRUE)] <- "DUST"
new_storm_data$EVTYPE<-factor(new_storm_data$EVTYPE)
```

Property damage and crop damage are each represented by two variables. The following code represents the damage as a value that can be easily compared acrossed EVTYPE.

```
new_storm_data$PROPDMGEXP<-recode(new_storm_data$PROPDMGEXP,'K'=1000,'M'=1000000,'B'=1000000000,.defaul
new_storm_data$CROPDMGEXP<-recode(new_storm_data$CROPDMGEXP,'K'=1000,'M'=1000000,'B'=1000000000,.defaul
new_storm_data$PROPDMGVALUE <- new_storm_data$PROPDMG*new_storm_data$PROPDMGEXP
new_storm_data$CROPDMGVALUE <- new_storm_data$CROPDMG*new_storm_data$CROPDMGEXP
```

Across the United States, which types of events (as indicated in the EVTYPE variable) are most harmful with respect to population health?

```
health<-(new_storm_data %>% group_by(EVTYPE) %>% summarise(FATALITIES = sum(FATALITIES),INJURIES = sum(
```

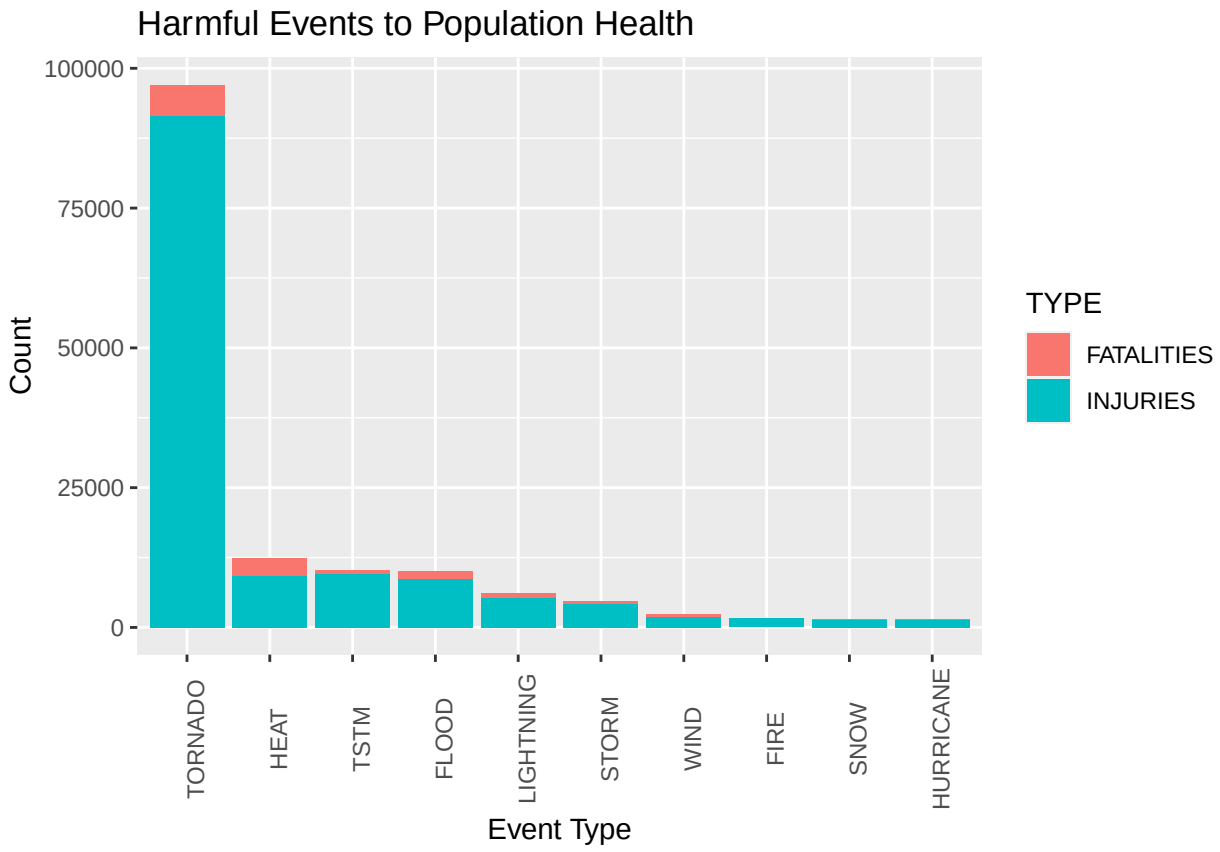
```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
most_harmful<-health[1:10,]
print(most_harmful)
```

```
## # A tibble: 10 x 3
##    EVTYPE    FATALITIES INJURIES
##    <fct>          <dbl>    <dbl>
##  1 TORNADO         5661    91407
##  2 HEAT            3178     9243
##  3 TSTM             729     9544
##  4 FLOOD           1525     8604
##  5 LIGHTNING        817     5231
##  6 STORM            423     4211
##  7 WIND             447     1890
##  8 FIRE              90     1608
##  9 SNOW             180     1331
```

```
## 10 HURRICANE        135     1328
```

```
plotdata<-gather(most_harmful, TYPE, VALUE, FATALITIES:INJURIES)
p <- ggplot(plotdata, aes(x=reorder(EVTYPE,-VALUE), y=VALUE, fill=TYPE))+geom_bar(stat="identity")+labs
p + theme(axis.text.x = element_text(angle = 90))
```



Across the United States, which types of events have the greatest economic consequences?

```
econ <-(new_storm_data %>% group_by(EVTYPE) %>% summarise(PROPDMGVALUE = sum(PROPDMGVALUE),CROPDMGVALUE
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
most_econ<-econ[1:10,]
print(most_econ)
```

```
## # A tibble: 10 x 3
##    EVTYPE     PROPDMGVALUE CROPDMGVALUE
##    <fct>             <dbl>        <dbl>
##  1 FLOOD     167529740932.  12380099110
##  2 HURRICANE  84736180030    5505292810
##  3 STORM      66624207661    5753913500
##  4 TORNADO    58581598040.    417461520
##  5 HAIL       15728664053.   3046471040
##  6 DROUGHT     1052838600   13972581000
##  7 TSTM       10973095481.   1271674992
##  8 FIRE        8496628500     403281630
##  9 WIND        6072889123     750310405
```

```
## 10 RAIN        3241445690     806162800
```

```
plotdata2<-gather(most_econ, TYPE, VALUE, PROPDMGVALUE:CROPDMGVALUE)
pe <- ggplot(plotdata2, aes(x=reorder(EVTYPE,-VALUE), y=VALUE, fill=factor(TYPE, labels=c("crop damage"
pe + theme(axis.text.x = element_text(angle = 90))
```