Project2_Ed_StormData

EdGomes

4 de novembro de 2017

Guide

1. *Synopsis*
2. *Loading and preprocessing the data Load the data (i.e. read.csv()) Process/transform the data (if necessary) into a format suitable for your analysis*

Decompressing files

Harmful Events to Population Health

Economic Consequences

3. *Results*

Harmful Events to Population Health

Economic Consequences

---

1. Synopsis

This project satisfies the Johns Hopkins Reproducible Research course offered through Coursera, Project 2 requirements. It loads a data set, performs some processing, and produces some charts. The data involved concerns storm and other severe weather events can cause both public health and economic problems.

This project involves exploring the U.S. National Oceanic and Atmospheric Administration's (NOAA) storm database. Many severe events can result in fatalities, injuries, and property damage, and preventing such outcomes to the extent possible is a key concern.

2.Loading and preprocessing the data

The data is in an unzipped file named "StormData.csv.bz2". If it has not already been loaded, it is loaded into memory.

```
ArqStorm <- "https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2 "
destStorm<-"D:/OneDrive/Software R/coursera/Reproducible Research/Project2/StormData.csv.bz2"
download.file(ArqStorm,destStorm, mode = "wb")
```

Decompressing files *StormData.csv.bz2*

```
if(!file.exists('StormData.csv')){
 library(R.utils)
  bunzip2("StormData.csv.bz2", "StormData.csv")}
```

Cleaning Storm Data

- *Loading and Exploring Data to first ideas of analysis*

```
Storm <- read.csv("StormData.csv")
names(Storm)
```

```
##  [1] "STATE__"    "BGN_DATE"   "BGN_TIME"   "TIME_ZONE"  "COUNTY"
##  [6] "COUNTYNAME" "STATE"      "EVTYPE"     "BGN_RANGE"  "BGN_AZI"
## [11] "BGN_LOCATI" "END_DATE"   "END_TIME"   "COUNTY_END" "COUNTYENDN"
## [16] "END_RANGE"  "END_AZI"    "END_LOCATI" "LENGTH"     "WIDTH"
## [21] "F"          "MAG"        "FATALITIES" "INJURIES"   "PROPDMG"
## [26] "PROPDMGEXP" "CROPDMG"    "CROPDMGEXP" "WFO"        "STATEOFFIC"
## [31] "ZONENAMES"  "LATITUDE"   "LONGITUDE"  "LATITUDE_E" "LONGITUDE_"
## [36] "REMARKS"    "REFNUM"
```

Harmful Events to Population Health

- *To evaluate harmful events to population Health, the fatalities and injuries total for each event type (EVTYPE) are calculated. The code chunk for this calculation are shown as follows.*

```
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
```

```
Storm_Fatalities <- Storm %>%
  select(EVTYPE, FATALITIES) %>%
  group_by(EVTYPE) %>%
  summarise(total.fatalities = sum(FATALITIES)) %>%
  arrange(-total.fatalities)

head(Storm_Fatalities, 20)
```

```
## # A tibble: 20 x 2
##                    EVTYPE total.fatalities
##                    <fctr>            <dbl>
##  1                TORNADO             5633
##  2         EXCESSIVE HEAT             1903
##  3            FLASH FLOOD              978
##  4                   HEAT              937
##  5              LIGHTNING              816
##  6              TSTM WIND              504
```

```
##  7                  FLOOD             470
##  8             RIP CURRENT            368
##  9               HIGH WIND            248
## 10               AVALANCHE            224
## 11            WINTER STORM            206
## 12            RIP CURRENTS            204
## 13               HEAT WAVE            172
## 14            EXTREME COLD            160
## 15       THUNDERSTORM WIND            133
## 16              HEAVY SNOW            127
## 17 EXTREME COLD/WIND CHILL            125
## 18             STRONG WIND            103
## 19                BLIZZARD            101
## 20               HIGH SURF            101
```

```
Storm_injuries <- Storm %>%
  select(EVTYPE, INJURIES) %>% group_by(EVTYPE) %>%
  summarise(total.injuries = sum(INJURIES)) %>%
  arrange(-total.injuries)

head(Storm_injuries, 20)
```

```
## # A tibble: 20 x 2
##                EVTYPE total.injuries
##                <fctr>          <dbl>
##  1             TORNADO          91346
##  2           TSTM WIND           6957
##  3               FLOOD           6789
##  4     EXCESSIVE HEAT           6525
##  5           LIGHTNING           5230
##  6                HEAT           2100
##  7           ICE STORM           1975
##  8         FLASH FLOOD           1777
##  9  THUNDERSTORM WIND           1488
## 10                HAIL           1361
## 11        WINTER STORM           1321
## 12  HURRICANE/TYPHOON           1275
## 13           HIGH WIND           1137
## 14          HEAVY SNOW           1021
## 15            WILDFIRE            911
## 16 THUNDERSTORM WINDS            908
## 17            BLIZZARD            805
## 18                 FOG            734
## 19   WILD/FOREST FIRE            545
## 20          DUST STORM            440
```

Economic Consequences

```
Storm.Economic <- Storm%>%
  select(EVTYPE, PROPDMG,PROPDMGEXP,CROPDMG,CROPDMGEXP)
```

```
Symbol <- sort(unique(as.character(Storm.Economic$PROPDMGEXP)))
Symbol
```

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

```
Mult <- c(0,0,0,1,10,10,10,10,10,10,10,10,10,10^9,10^2,10^2,10^3,10^6,10^6)
```

- *PROPDMGEXP and CROPDMGEXP can be interpreted as the following:*

"*blank*" -> x 0

"-" -> x 0

"?" -> x 0

"+" -> x 1

"H", "h" -> hundreds = x 100

"K", "K" -> kilos = x 1,000

"M", "m" -> millions = x 1,000,000

"B","b" -> billions = x 1,000,000,000

```
Convert <- data.frame(Symbol, Mult)
Convert
```

```
##    Symbol  Mult
## 1         0e+00
## 2       - 0e+00
## 3       ? 0e+00
## 4       + 1e+00
## 5       0 1e+01
## 6       1 1e+01
## 7       2 1e+01
## 8       3 1e+01
## 9       4 1e+01
## 10      5 1e+01
## 11      6 1e+01
## 12      7 1e+01
## 13      8 1e+01
## 14      B 1e+09
## 15      h 1e+02
## 16      H 1e+02
## 17      K 1e+03
## 18      m 1e+06
## 19      M 1e+06
```

```
Storm.Economic$PropMult<- Convert$Mult[match(Storm$PROPDMGEXP, Convert$Symbol)]
Storm.Economic$CropMult<- Convert$Mult[match(Storm$CROPDMGEXP, Convert$Symbol)]
```

```
Storm.Economic<- Storm.Economic %>%
  mutate(PROPDMG = PROPDMG*PropMult) %>%
  mutate(CROPDMG = CROPDMG*CropMult) %>%
  mutate(TOTAL.DMG = PROPDMG+CROPDMG)

Economic.total <- Storm.Economic %>%
  group_by(EVTYPE) %>%
  summarize(TOTAL.DMG.EVTYPE = sum(TOTAL.DMG))%>%
  arrange(-TOTAL.DMG.EVTYPE)
```

```
head(Storm.Economic,20)
```

```
##     EVTYPE PROPDMG PROPDMGEXP CROPDMG CROPDMGEXP PropMult CropMult
## 1  TORNADO   25000          K       0                1e+03        0
## 2  TORNADO    2500          K       0                1e+03        0
## 3  TORNADO   25000          K       0                1e+03        0
## 4  TORNADO    2500          K       0                1e+03        0
## 5  TORNADO    2500          K       0                1e+03        0
## 6  TORNADO    2500          K       0                1e+03        0
## 7  TORNADO    2500          K       0                1e+03        0
## 8  TORNADO    2500          K       0                1e+03        0
## 9  TORNADO   25000          K       0                1e+03        0
## 10 TORNADO   25000          K       0                1e+03        0
## 11 TORNADO 2500000          M       0                1e+06        0
## 12 TORNADO 2500000          M       0                1e+06        0
## 13 TORNADO  250000          K       0                1e+03        0
## 14 TORNADO       0          K       0                1e+03        0
## 15 TORNADO   25000          K       0                1e+03        0
## 16 TORNADO   25000          K       0                1e+03        0
## 17 TORNADO   25000          K       0                1e+03        0
## 18 TORNADO   25000          K       0                1e+03        0
## 19 TORNADO   25000          K       0                1e+03        0
## 20 TORNADO   25000          K       0                1e+03        0
##    TOTAL.DMG
## 1      25000
## 2       2500
## 3      25000
## 4       2500
## 5       2500
## 6       2500
## 7       2500
## 8       2500
## 9      25000
## 10     25000
## 11   2500000
## 12   2500000
## 13    250000
## 14         0
## 15     25000
## 16     25000
## 17     25000
## 18     25000
## 19     25000
## 20     25000
```

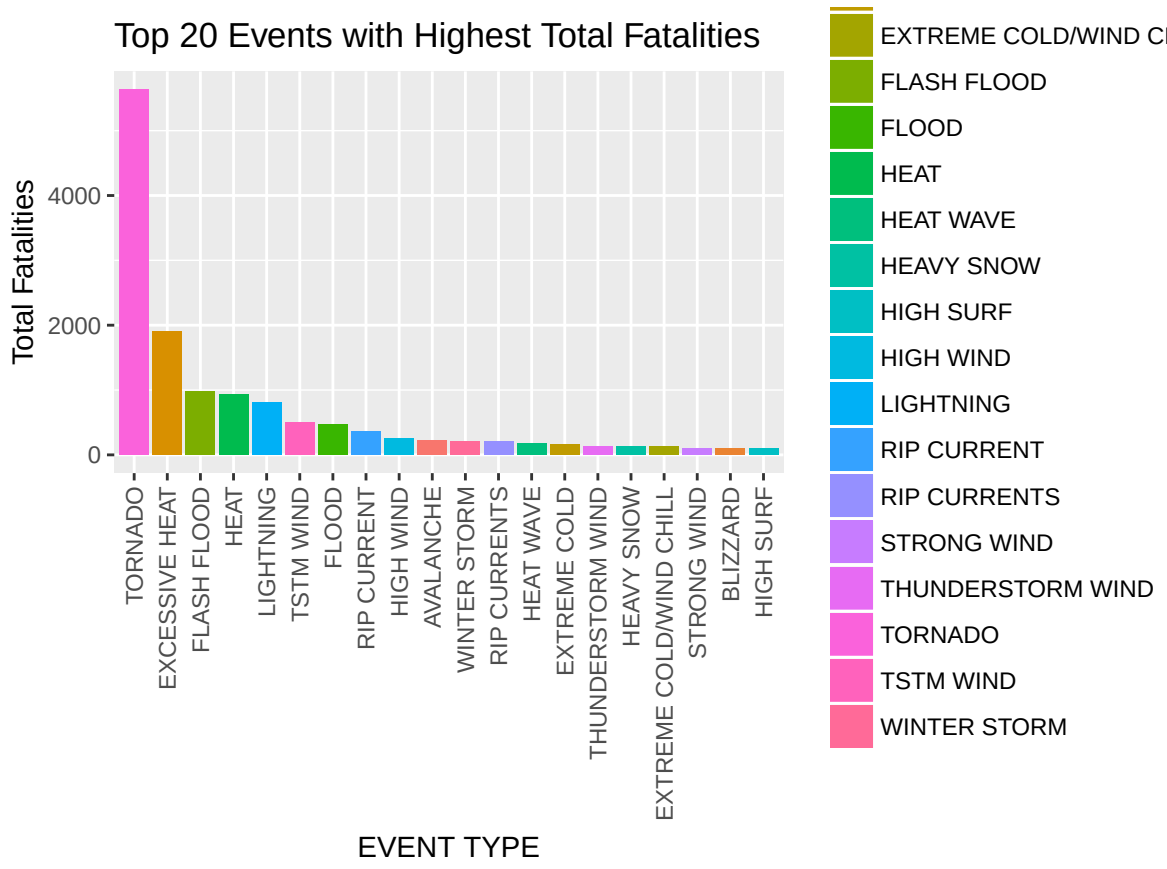
---

3. Results

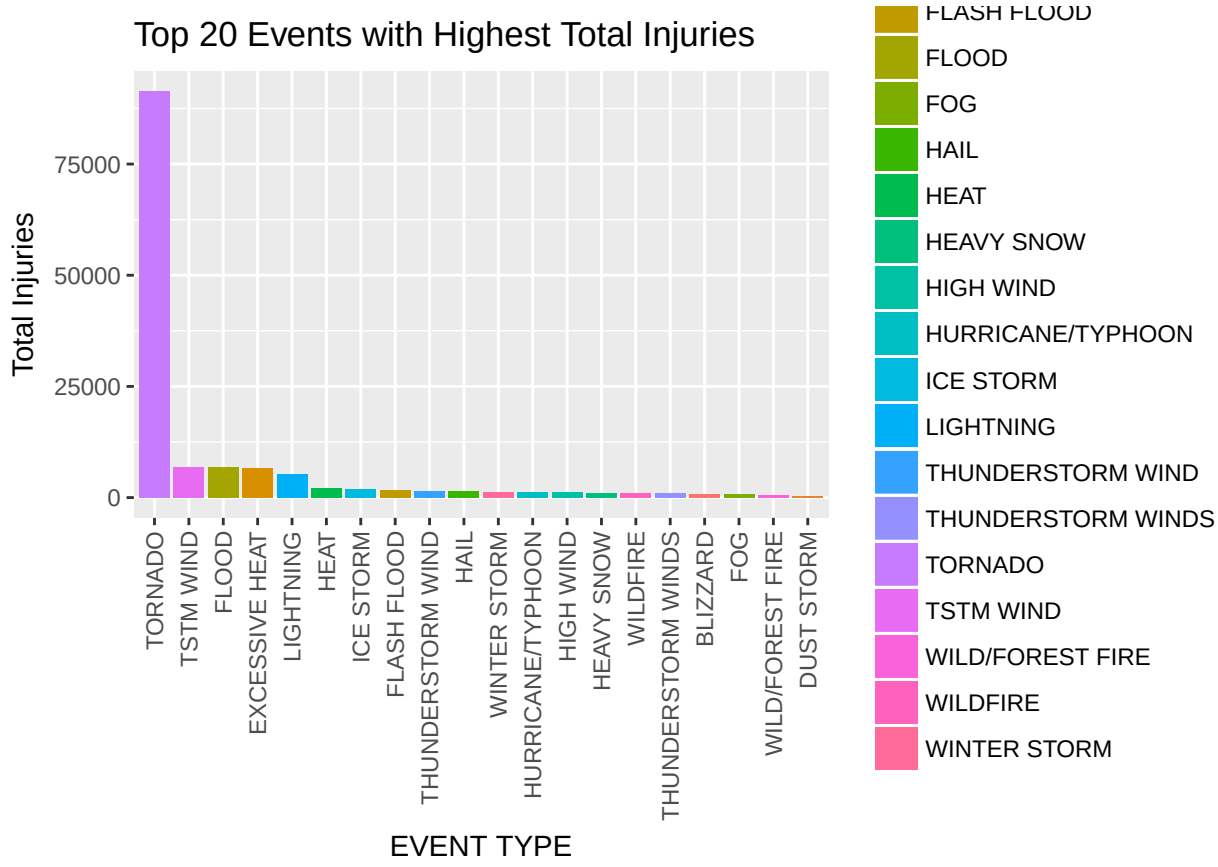
Harmful Events to Population Health

The top 20 events with the highest total fatalities and injuries are shown graphically.

```
  library(ggplot2)
Figure1 <- ggplot(Storm_Fatalities[1:20,], aes(x=reorder(EVTYPE, -total.fatalities), y=total.fatalities
print(Figure1)
```



```
Figure2 <- ggplot(Storm_injuries[1:20,], aes(x=reorder(EVTYPE, -total.injuries), y=total.injuries, fill=
print(Figure2)
```



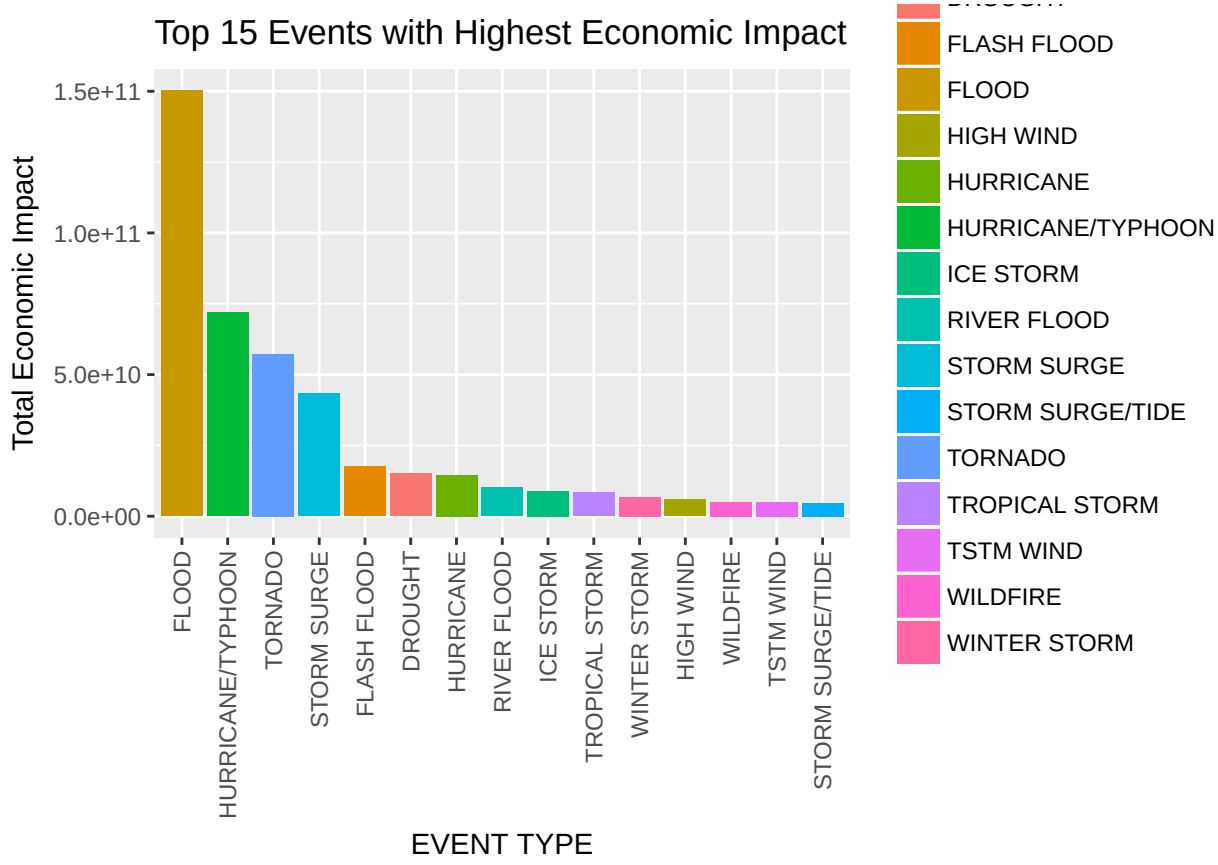
Across the United States, which types of events (as indicated in the EVTYPE variable) are most harmful with respect to population health?

*Event type with the most harmful weather to population health was "**TORNADO**"*

Economic Consequences

Top 15 weather event types have the greatest economic consequences are shown graphically.

```
Figure3<- ggplot(Economic.total[1:15,], aes(x=reorder(EVTYPE, -TOTAL.DMG.EVTYPE), y=TOTAL.DMG.EVTYPE, f
print(Figure3)
```



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

*Event type with worst economic consequence was "**FLOOD**".*