Week4-Project2

Synopsis

Storms and other severe weather events can cause both public health and economic problems for communities and municipalities. Many severe events can result in fatalities, injuries, and property damage, and preventing such outcomes to the extent possible is a key concern.

This project involves exploring the U.S. National Oceanic and Atmospheric Administration's (NOAA) storm database. This database tracks characteristics of major storms and weather events in the United States, including when and where they occur, as well as estimates of any fatalities, injuries, and property damage.

Data Processing

```r
data <- read.csv("StormData.csv")
head(data)
```

```
##   STATE__           BGN_DATE BGN_TIME TIME_ZONE COUNTY COUNTYNAME STATE
## 1       1  4/18/1950 0:00:00     0130       CST     97     MOBILE    AL
## 2       1  4/18/1950 0:00:00     0145       CST      3    BALDWIN    AL
## 3       1  2/20/1951 0:00:00     1600       CST     57    FAYETTE    AL
## 4       1   6/8/1951 0:00:00     0900       CST     89    MADISON    AL
## 5       1 11/15/1951 0:00:00     1500       CST     43    CULLMAN    AL
## 6       1 11/15/1951 0:00:00     2000       CST     77 LAUDERDALE    AL
##    EVTYPE BGN_RANGE BGN_AZI BGN_LOCATI END_DATE END_TIME COUNTY_END
## 1 TORNADO         0                                               0
## 2 TORNADO         0                                               0
## 3 TORNADO         0                                               0
## 4 TORNADO         0                                               0
## 5 TORNADO         0                                               0
## 6 TORNADO         0                                               0
##   COUNTYENDN END_RANGE END_AZI END_LOCATI LENGTH WIDTH F MAG FATALITIES
## 1         NA         0                      14.0   100 3   0          0
## 2         NA         0                       2.0   150 2   0          0
## 3         NA         0                       0.1   123 2   0          0
## 4         NA         0                       0.0   100 2   0          0
## 5         NA         0                       0.0   150 2   0          0
## 6         NA         0                       1.5   177 2   0          0
##   INJURIES PROPDMG PROPDMGEXP CROPDMG CROPDMGEXP WFO STATEOFFIC ZONENAMES
## 1       15    25.0          K       0
## 2        0     2.5          K       0
## 3        2    25.0          K       0
## 4        2     2.5          K       0
## 5        2     2.5          K       0
## 6        6     2.5          K       0
##   LATITUDE LONGITUDE LATITUDE_E LONGITUDE_ REMARKS REFNUM
## 1     3040      8812       3051       8806              1
## 2     3042      8755          0          0              2
## 3     3340      8742          0          0              3
## 4     3458      8626          0          0              4
## 5     3412      8642          0          0              5
## 6     3450      8748          0          0              6
```

```r
names(data)
```

```
##  [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"
```

Let's pick the useful information columns.

```r
data.h <- data[, c(8,23:24)]
data.p <- data[, c(8, 25:28)]
```

Result

1. Find the most harmful type of event with respect to population health.

In this case, we discover from the number of fatalities and injuries.

```r
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
```

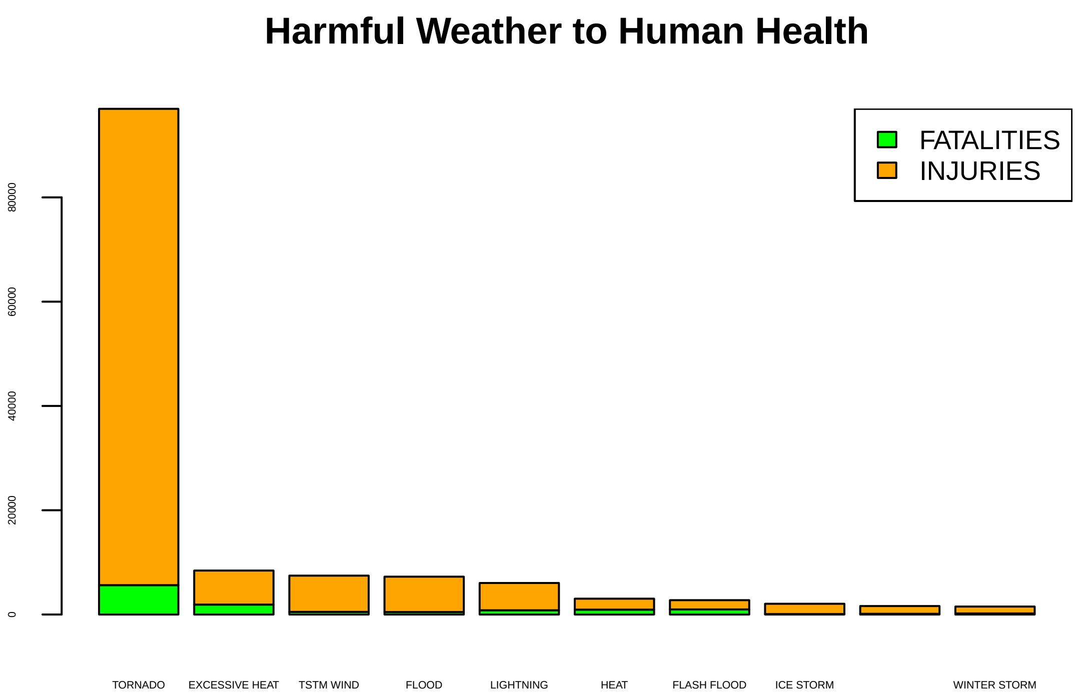
```r
data.h2 <- aggregate(cbind(FATALITIES,INJURIES) ~ EVTYPE, data = data.h, sum, na.rm = TRUE)
rank.h <- arrange(data.h2, desc(FATALITIES+INJURIES))
top10.h <- rank.h[1:10,]
top10.h
```

```
##               EVTYPE FATALITIES INJURIES
## 1            TORNADO       5633    91346
## 2     EXCESSIVE HEAT       1903     6525
## 3          TSTM WIND        504     6957
## 4              FLOOD        470     6789
## 5          LIGHTNING        816     5230
## 6               HEAT        937     2100
## 7        FLASH FLOOD        978     1777
## 8          ICE STORM         89     1975
## 9  THUNDERSTORM WIND        133     1488
## 10      WINTER STORM        206     1321
```

The following is a graphical way to look at the most harmful weather event to human health.

```r
FAT_INJ <- as.matrix(t(top10.h[,-1]))
colnames(FAT_INJ) <- top10.h$EVTYPE
```

```r
barplot(FAT_INJ, col=c("green","orange"), cex.names = 0.3, cex.axis = 0.3, main = "Harmful Weather to Hu
legend("topright", c("FATALITIES", "INJURIES"), fill=c("green", "orange"), cex=0.8)
```



2. Find the type of event has the greates economic consequences.

Let's start from converting different scale units to same scale.

```r
unique(data.p$PROPDMGEXP)
```

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

```r
unique(data.p$CROPDMGEXP)
```

```
## [1]   M K m B ? 0 k 2
## Levels:  ? 0 2 B k K m M
```

```r
table(data.p$PROPDMGEXP)
```

```
##
##              -      ?      +      0      1      2      3      4      5
## 465934       1      8      5    216     25     13      4      4     28
##      6       7      8      B      h      H      K      m      M
##      4       5      1     40      1      6 424665      7  11330
```

```r
table(data.p$CROPDMGEXP)
```

```
##
##              ?      0      2      B      k      K      m      M
## 618413       7     19      1      9     21 281832      1   1994
```

Let's tune the data first, assign new factor levels and define the NA values in the following way.

```r
data.p$prop <- factor(data.p$PROPDMGEXP, levels=c("H", "K", "M", "B", "h", "m", "O"))
data.p$prop[is.na(data.p$prop)] <- "O"
table(data.p$prop)
```

```
##
##      H      K      M      B      h      m      0
##      6 424665  11330     40      1      7 466248
```

```r
# double check, so we get rid of other symbols and missing values

data.p$crop <- factor(data.p$CROPDMGEXP, levels=c("K", "M", "B", "k", "m", "0"))
data.p$crop[is.na(data.p$crop)] <- "0"
table(data.p$crop)
```

```
##
##      K      M      B      k      m      0
## 281832   1994      9     21      1 618440
```

Now let's assign numerical values to change unit symbols.

```r
data.p$PROP[data.p$prop == "K"] <- 1000
data.p$PROP[data.p$prop == "H" | data.p$prop == "h"] <- 100
data.p$PROP[data.p$prop == "M" | data.p$prop == "m"] <- 1000000
data.p$PROP[data.p$prop == "B"] <- 1000000000
data.p$PROP[data.p$prop == "0"] <- 1

data.p$CROP[data.p$crop == "K" | data.p$crop == "k"] <- 100
data.p$CROP[data.p$crop == "M" | data.p$crop == "m"] <- 1000000
data.p$CROP[data.p$crop == "B"] <- 1000000000
data.p$CROP[data.p$crop == "0"] <- 1
```

Create two new columns of values based on the standard units.

```r
data.p <- mutate(data.p, prop_value = PROPDMG*PROP/1000000, crop_value = CROPDMG*CROP/1000000)
data.p2 <- aggregate(cbind(prop_value, crop_value) ~ EVTYPE, data = data.p, sum, na.rm = TRUE)
data.p2 <- data.p2 %>% group_by(EVTYPE) %>% summarize(prop_value = sum(prop_value, na.rm = TRUE), crop_
data.p2 <- arrange(data.p2, desc(prop_value+crop_value))
top10.p <- data.p2[1:10,]
top10.p
```

```
## # A tibble: 10 x 3
##    EVTYPE            prop_value   crop_value
##    <fct>                  <dbl>        <dbl>
##  1 FLOOD                144658.  5516.
##  2 HURRICANE/TYPHOON     69306.  2605.
##  3 TORNADO               56937.   325.
##  4 STORM SURGE           43324.     0.000500
##  5 HAIL                  15732.  2507.
##  6 FLASH FLOOD           16141.  1261.
##  7 DROUGHT                1046. 13953.
##  8 HURRICANE             11868.  2740.
##  9 RIVER FLOOD            5119.  5026.
## 10 ICE STORM              3945.  5021.
```

The following is a graphical way to look at the most harmful weather event to economic impact.

```r
PROP_CROP <- as.matrix(t(top10.p[,-1]))
colnames(PROP_CROP) <- top10.p$EVTYPE
barplot(PROP_CROP, col=c("green","orange"), cex.names = 0.3, cex.axis = 0.3, main = "Harmful Weather to
legend("topright", c("Property", "Crop"), fill=c("green", "orange"), cex=0.8)
```

Harmful Weather to Economic