Weather Season Synopsis

YASIR MUNIR

6/3/2020

Synopsis

NOAA (National Oceanic & Atmospheric Administration) is the world's largest provider of weather and climate data. In particular with severe weather events such as blizzards and tornadoes, the Storm Events Database collects information on significant weather phenomena, dating back to January 1950. The following report aims to answer questions on the types of events that have caused the greatest health and economic impact since such time.

Data Processing

Note: According to the Storm Events Database Details Link; From Jan 1950 to Dec 1995, only Tornado, Thunderstorm Wind and Hail had been recorded as Severe weather events. As such, to depict a true picture of impact for all 48 event types, the dataset will be from 1996 to 2011.

Required R Libraries

```
library(data.table)library(stringr)library(dplyr)
```

Read Storm Data and Process

```
csv_name <- "repdata_data_StormData.csv.bz2"

# Read file

storm_data <-fread(csv_name)

# Format date variables

storm_data <- storm_data[,BDate := as.POSIXct(word(BGN_DATE,1), format = "%m/%d/%Y")]

# Reduce dataset to 1996 onwards

storm_data <- storm_data[format(BDate, "%Y")> 1995,]
```

Analyze Data Structure

```
str(storm_data)

## Classes 'data.table' and 'data.frame':   653530 obs. of  38 variables:
```

```
##  $ STATE__   : num  1 1 1 1 1 1 1 1 1 1 ...
##  $ BGN_DATE : chr  "1/6/1996 0:00:00" "1/11/1996 0:00:00" "1/11/1996
0:00:00" "1/11/1996 0:00:00" ...
##  $ BGN_TIME  : chr  "08:00:00 PM" "06:35:00 PM" "06:45:00 PM" "07:05:00
PM" ...
##  $ TIME_ZONE : chr  "CST" "CST" "CST" "CST" ...
##  $ COUNTY    : num  1 31 31 45 67 125 1 75 51 101 ...
##  $ COUNTYNAME: chr  "ALZ001>038" "COFFEE" "COFFEE" "DALE" ...
##  $ STATE     : chr  "AL" "AL" "AL" "AL" ...
##  $ EVTYPE    : chr  "WINTER STORM" "TORNADO" "TSTM WIND" "TSTM WIND"
 ...
##  $ BGN_RANGE : num  0 5 0 0 8 0 0 8 23 ...
##  $ BGN_AZI   : chr  "" "N" "" "" ...
##  $ BGN_LOCATI: chr  "" "KINSTON" "ENTERPRISE" "PINCKARD" ...
##  $ END_DATE  : chr  "1/7/1996 0:00:00" "1/11/1996 0:00:00" "1/11/1996
0:00:00" "1/11/1996 0:00:00" ...
##  $ END_TIME  : chr  "03:00:00 PM" "06:36:00 PM" "06:45:00 PM" "07:05:00
PM" ...
##  $ COUNTY_END: num  0 0 0 0 0 0 0 0 0 ...
##  $ COUNTYENDN: logi  NA NA NA NA NA NA ...
##  $ END_RANGE : num  0 5 0 0 8 0 0 8 0 ...
##  $ END_AZI   : chr  "" "N" "" "" ...
##  $ END_LOCATI: chr  "" "KINSTON" "ENTERPRISE" "PINCKARD" ...
##  $ LENGTH    : num  0 1 0 0 0 0 0 0 0 ...
##  $ WIDTH     : num  0 75 0 0 0 0 0 0 0 ...
##  $ F         : int  NA 1 NA NA NA NA NA NA NA NA ...
##  $ MAG       : num  0 0 0 0 75 40 50 50 50 ...
##  $ FATALITIES: num  0 0 0 0 0 0 0 0 0 ...
##  $ INJURIES  : num  0 0 0 0 0 0 0 0 0 ...
##  $ PROPDMG   : num  380 100 3 5 2 0 400 12 8 12 ...
```

```
##  $ PROPDMGEXP: chr  "K" "K" "K" "K" ...
##  $ CROPDMG   : num  38 0 0 0 0 0 0 0 0 ...
##  $ CROPDMGEXP: chr  "K" "" "" "" ...
##  $ WFO       : chr  "BMX" "TAE" "TAE" "TAE" ...
##  $ STATEOFFIC: chr  "ALABAMA, Central" "ALABAMA, Southeast" "ALABAMA, So
utheast" "ALABAMA, Southeast" ...
##  $ ZONENAMES : chr  "LAUDERDALE - LAUDERDALE - COLBERT - FRANKLIN
 - LAWRENCE - LIMESTONE - MADISON - MORGAN - MARSHALL - JACKSON -
D"| _truncated_ "" "" "" ...
##  $ LATITUDE  : num  0 3116 3119 3119 3121 ...
##  $ LONGITUDE : num  0 8608 8551 8533 8521 ...
##  $ LATITUDE_E: num  0 3116 3119 3119 3121 ...
##  $ LONGITUDE_: num  0 8608 8551 8533 8521 ...
##  $ REMARKS   : chr  "A winter storm brought a mixture of freezing rain, sleet,
and snow to the northern two-thirds of Alabama.  Prec"| _truncated_ "A tornado d
estroyed 4 house trailers that were unoccupied. Debris was scattered for about 1 mile,
 according to "| _truncated_ "Several trees were blown down and two backyard she
ds were destroyed according to newspaper reports and county e"| _truncated_ "Coun
ty emergency management confirmed that three sheds were destroyed, and several h
ouses received superficial damage." ...
##  $ REFNUM    : num  248768 248769 248770 248771 248772 ...
##  $ BDate     : POSIXct, format: "1996-01-06" "1996-01-11" ...
##  - attr(*, ".internal.selfref")=<externalptr>
```

Review Property Damage Exponents

```
unique(storm_data$PROPDMGEXP)
```

```
## [1] "K" ""  "M" "B" "0"
```

Review Crop Damage Exponents

```
unique(storm_data$CROPDMGEXP)
```

```
## [1] "K" ""  "M" "B"
```

Translate Damage Exponents

```
# PROP CONVERT ALL CHARACTERS TO A NUMBER EXPONENT

storm_data$PROPDMG2 <- ifelse(storm_data$PROPDMGEXP=="B", 9,

                              ifelse(storm_data$PROPDMGEXP=="M", 6,

                                     ifelse(storm_data$PROPDMGEXP=="K", 6,

                                            ifelse(storm_data$PROPDMGEXP=="", 0,storm_data$PROPDMGEXP))))


# CROP CONVERT ALL CHARACTERS TO A NUMBER EXPONENT

storm_data$CROPDMG2 <- ifelse(storm_data$CROPDMGEXP=="B", 9,

                              ifelse(storm_data$CROPDMGEXP=="M", 6,

                                     ifelse(storm_data$CROPDMGEXP=="K", 3,

                                            ifelse(storm_data$CROPDMGEXP=="-", 0,

                                                   ifelse(storm_data$CROPDMGEXP=="", 0,

                                                          ifelse(storm_data$CROPDMGEXP=="?", 0,storm_data$CROPDMGEXP))))))


# Tally Damage from property and crops

storm_data$TotalDamage <- storm_data$PROPDMG*10^(as.numeric(storm_data$PROPDMG2)) + storm_data$CROPDMG*10^(as.numeric(storm_data$CROPDMG2))
```

Results

Question 1

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

```
# Group By Fatalities

data_agg_deaths <- storm_data %>%

mutate(Year = format(BDate, "%Y")) %>% group_by(EVTYPE) %>% summarise(total = sum(FATALITIES))

# Arrange and Remove zero totals

data_agg_deaths <- data_agg_deaths[data_agg_deaths$total > 0,]

data_agg_deaths <- data_agg_deaths %>% arrange(desc(total))

# EVtypes from top 5% of total deaths between 1996-2011

top_data_agg_deaths <- data_agg_deaths[data_agg_deaths$total > quantile(data_agg_deaths$total, prob=1-5/100),]


colnames(top_data_agg_deaths) <- c("EventType", "Total")

# Group By Injuries

data_agg_injured <- storm_data %>%

mutate(Year = format(BDate, "%Y")) %>% group_by(EVTYPE) %>% summarise(total = sum(INJURIES))

# Arrange and Remove zero totals

data_agg_injured <- data_agg_injured[data_agg_injured$total > 0,]

data_agg_injured <- data_agg_injured %>% arrange(desc(total))

# EVtypes from top 5% of total injuries between 1996-2011

top_data_agg_injured <- data_agg_injured[data_agg_injured$total > quantile(data_agg_injured$total, prob=1-5/100),]


colnames(top_data_agg_injured) <- c("EventType", "Total")
```

Plot Fatalities by EV Types

```
barplot(table(top_data_agg_deaths$EventType), height=top_data_agg_deaths$Total, names=top_data_agg_deaths$EventType, col=rainbow(length(top_data_agg_deaths$EventType)), cex.name=0.5, cex.axis = 0.75, las=2, horiz=T, main="U.S. Fatalities by Event Type Between 1996-2011")
```

Plot Injuries by EV Types

```
barplot(table(top_data_agg_injured$EventType), height=top_data_agg_injured$Total, names=top_data_agg_injured$EventType, col=rainbow(length(top_data_agg_injured$EventType)), cex.name=0.5, cex.axis = 0.75, las=2, horiz=T, main="U.S. Injuries by Event Type Between 1996-2011")
```

Question 2

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

```
# Group By Damages

data_agg_damages <- storm_data %>%

mutate(Year = format(BDate, "%Y")) %>% group_by(EVTYPE) %>% summarise(total = sum(TotalDamage))
```

```
# Arrange and Remove zero totals

data_agg_damages <- data_agg_damages[data_agg_damages$total > 0,]

data_agg_damages <- data_agg_damages %>% arrange(desc(total))

# EVtypes from top 5% of total damages between 1996-2011

top_data_agg_damages <- data_agg_damages[data_agg_damages$total > quantile(data_agg_damages$total, prob=1-5/100),]




top_data_agg_damages$total <- top_data_agg_damages$total/1000000000



top_data_agg_damages$total <- format(round(top_data_agg_damages$total,2), nsmall=2)



top_data_agg_damages$total <- as.numeric(top_data_agg_damages$total)



colnames(top_data_agg_damages) <- c("EventType", "USD_BILLIONS")
```

Plot Economic Impact ($)

```
barplot(table(top_data_agg_damages$EventType), height=top_data_agg_damages$USD_BILLIONS, names=top_data_agg_damages$EventType, col=rainbow(length(top_data_agg_damages$EventType)), cex.name=0.5, cex.axis = 0.75, las=2, horiz=T, main="Economic Impact by Event Type Between 1996-2011 (In USD Billions)")
```